Identification of Cellular Host Factors That Associate With LINE-1 ORF1p and the Effect of the Zinc Finger Antiviral Protein ZAP on LINE-1 Retrotransposition

by

John B. Moldovan

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Cellular and Molecular Biology)
in the University of Michigan
2015

Doctoral Committee:

Professor John V. Moran, Chair
Professor David R. Engelke
Professor Michael J. Imperiale
Professor Diane M. Robins
Associate Professor JoAnn Sekiguchi

To my family.

Acknowledgements

First I would like to thank my mentor, Dr. John Moran. John's integrity and scientific rigor are matched equally by his infectious enthusiasm and dedication to training students and to the people in his charge. John pushes his students beyond their perceived limits...and then just a little bit more. Importantly, John maintains perspective and knows when to say, "Just have fun with it."

I also would like to thank Dr. Dave Engelke, Dr. Mike Imperiale, Dr. Didi Robins, and Dr. JoAnn Sekiguchi for serving on my thesis committee. They asked tough questions and provided critical feedback, which significantly influenced the final shape and direction of my dissertation.

I would also like to thank all current and former Moran lab members. They are a most thoughtful, intelligent, imaginative, somewhat eccentric, sometimes critical, and always supportive bunch of characters that made the lab a fun, enjoyable, and very interesting place to work in. I specifically need to thank Darko Plavcic, an undergraduate student (at the time) who helped with the many, many retrotransposition assays that needed to be done during the course of my research.

Finally, I want to thank my family and friends for their unwavering and unconditional support throughout - past, present, and beyond...

Table of Contents

List of Figures

List of Appendices

Abstract

Identification of Cellular Host Factors That Associate With LINE-1 ORF1p and the Effect of the Zinc Finger Antiviral Protein ZAP on LINE-1 Retrotransposition

By

John B. Moldovan

Chair: John V. Moran

Long INterspersed Element-1 (LINE-1 or L1) is the only active autonomous retrotransposon in the human genome. The human genome contains over 500,000 L1 sequences, which account for approximately 17% of human DNA. L1 sequences mobilize throughout the human genome by a copy-and-paste mechanism known as retrotransposition. Most genomic L1 sequences are incapable of mobility (*i.e.*, retrotransposition) because they are either 5'-truncated, internally rearranged, and/or mutated; however, it is estimated that each human cell contains at least 80-100 intact L1 sequences that are retrotransposition capable.

L1 retrotransposition is inherently mutagenic and on occasion can disrupt gene expression leading to diseases such as hemophilia A and cancer. Due to the mutagenic potential of L1 retrotransposition, it thus stands to reason that the host cell has evolved mechanisms to protect the cell from unabated retrotransposition. In this thesis I identified cellular host factors that associate with the first L1 open reading frame protein, ORF1p. I demonstrate that the zinc finger antiviral protein ZAP associates with L1 ORF1p and inhibits human L1 and Alu retrotransposition as well as the retrotransposition of LINE elements from mice and zebrafish. Molecular genetic, biochemical, and fluorescence microscopy data suggest that ZAP interacts with L1 RNA and reduces the

expression of full-length L1 RNA and the L1-encoded proteins, thereby providing mechanistic insight into how ZAP may restrict retrotransposition.

In addition to ZAP, I show that the ORF1p-associated cellular host factors MOV10, hnRNPL, and PAR-4 also inhibit L1 retrotransposition. Mechanistic data suggest that ZAP, MOV10, hnRNPL, and PAR-4 restrict L1 retrotransposition by distinct mechanisms, suggesting that each of these cellular host factors may target different post-transcriptional steps in the L1 retrotransposition cycle.

Importantly, ZAP and MOV10 were first characterized as antiviral proteins due to their ability to suppress retroviral activity. Notably, several other host cell antiviral factors such as APOBEC3 proteins, TREX1, SAMHD1 and RNase L have recently been demonstrated to inhibit L1 retrotransposition. Thus, these data suggest that ZAP, MOV10 and perhaps other ORF1p-associated cellular host factors initially may have evolved to combat L1 and other endogenous retrotransposons and subsequently were co-opted as viral restriction factors.

Chapter 1

Introduction

Thesis Overview

The focus of my thesis addresses the question, what host cell proteins are involved in L1 retrotransposition? My thesis research began with a short study of the L1 retrotransposition mechanism in which I studied how mutations in the L1 ORF1p protein affect the function of L1 ribonucleoprotein particles (RNPs), an important L1 retrotransposition intermediate. The study of L1 RNPs then prompted me to question whether other cellular proteins might associate with L1 RNPs and modulate L1 retrotransposition activity. The following chapters will present the major results of my thesis research in the following order. Chapter 1 is an introductory chapter that provides background information on the mechanism of L1 retrotransposition, how L1 retrotransposition impacts the human genome, and what host cell processes are known to regulate L1 retrotransposition. Chapter 2 presents experiments that examine the role of ORF1p in L1 retrotransposition and how mutations in ORF1p affect the function of L1 RNPs. Chapter 3 describes the bulk of my thesis research where I used co-immunoprecipitation experiments in conjunction with tandem mass spectrometry to identify host cell proteins that associate with L1 RNPs, and then showed that some of these proteins may inhibit L1 retrotransposition. These experiments led to the discovery that the zinc-finger antiviral protein (ZAP) restricts human L1 and Alu retrotransposition. Chapter 4 extends findings from Chapter 3 and presents experiments that show that the RNA helicase MOV10, an additional L1 RNP-interacting protein identified in Chapter 3, also restricts L1 retrotransposition. In the final Chapter (Chapter 5), I discuss the significance of these findings, and focus on remaining questions and possible future directions.

Introduction to Transposable Elements

Transposable elements, sometimes referred to as “jumping genes,” are DNA sequences that can move from one DNA site to another within genomic DNA. Transposable elements were first discovered in maize during the 1940's by Barbara McClintock. She demonstrated that transposable elements could move to different chromosomal locations and cause heritable mutations that were linked to color variegation (Figure 1.1) (McClintock 1950, McClintock 1951).

Since their initial discovery, we have learned that transposable elements inhabit the genomes of virtually all organisms from each of the four kingdoms of life. Despite the ubiquity of transposable elements, why they exist and what function they might serve are puzzling and difficult questions to answer. Indeed, for many decades following their discovery transposable elements were considered a type of "junk DNA" (Ohno 1972), mainly because they do not appear to serve an obvious biological function for the organisms which they inhabit. Transposable elements also have been characterized as "selfish DNA" with their only purpose thought to be to spread more copies of themselves throughout an organisms’ genome (Doolittle and Sapienza 1980, Orgel and Crick 1980). Regardless of whether one thinks of transposable elements as pieces of "junk" or as inherently "selfish" entities, research within the last twenty years has revealed that transposable elements make a significant impact on the genomes and phenotypes of their hosts (Cordaux and Batzer 2009, Beck et al. 2011, Richardson et al. 2015). Whether transposable elements have an ultimate function or purpose may never be known, but it is clear that they are more than just junk. The continued study of transposable elements will provide insights into how they influence their host genomes, which will ultimately contribute to a more unified and comprehensive understanding of the basic biological processes that govern the function of single cells and entire organisms.

This introductory chapter will provide a brief overview of transposable elements in the human genome and then focus on the biology of the human long

interspersed nuclear element-1 (LINE-1 or L1), the only self-autonomous transposable element currently active in the human genome.

Transposable Elements in the Human Genome

The initial sequencing of the human genome revealed that nearly 50% of human DNA is composed of transposable elements (Lander et al. 2001) and more recent studies suggest that as much as 66%-69% of the human genome could be composed of transposable elements and other repetitive DNA sequences (de Koning et al. 2011). Transposable elements can be grouped into two main classes based on whether they mobilize via a DNA intermediate (DNA transposons) or an RNA intermediate (retrotransposons). The human genome contains a variety both DNA transposons and retrotransposons (Figure 1.2), which will be discussed below.

DNA transposons

DNA transposons comprise roughly 3% of the human genome. DNA transposons mobilize (*i.e.*, transpose) via a DNA intermediate by a cut-and-paste mechanism termed transposition. A classical DNA transposon typically consists of a pair of variable length terminal inverted repeat (TIR) sequences that surround a transposon-encoded enzyme called transposase (Figure 1.3A), which mediates the transposition reaction. During DNA transposition, transposase recognizes the TIR sequences on either end of the DNA transposon and catalyzes the excision and subsequent insertion of the DNA transposon from one place in DNA to another (Craig et al. 2002).

DNA transposons are not thought to be active in the human genome, and likely have been extinct (*i.e.*, not capable of transposition) in mammals for at least 37 million years (Lander et al. 2001, Pace and Feschotte 2007). Despite these facts, recent evidence suggests that a small group of DNA transposons may be active in certain bat species (Ray et al. 2007, Mitra et al. 2013). Although DNA transposons may not be currently active in humans, they appear to have left

an indelible mark on the human genome as several human proteins are thought to have evolved from DNA transposons including the RAG1 and RAG2 V(D)J recombinase enzymes, which function during B and T cell development (Thompson 1995, Lander et al. 2001, Kapitonov and Jurka 2005).

Retrotransposons

Retrotransposons are the most numerous class of transposable elements in the human genome (~42% of the genome) (Figure 1.2) (Lander et al. 2001). Retrotransposons mobilize via an RNA intermediate by a replicative copy-and-paste mechanism termed retrotransposition (Boeke et al. 1985, Craig et al. 2002, Beauregard et al. 2008, Cordaux and Batzer 2009). Retrotransposons typically encode a reverse transcriptase (RT) enzyme, which is required to mediate retrotransposition. During retrotransposition, an RNA copy of the retrotransposon is reverse transcribed by the retrotransposon-encoded RT to generate a retrotransposon cDNA copy, which is inserted into a new genomic location. The process of retrotransposition typically results in a net increase of +1 in retrotransposon copy number. In principle, the number of retrotransposon copies can increase over successive generations to comprise a significant portion of genomic DNA in some organisms, such as humans. All retrotransposons can be further subdivided into two main groups: long terminal repeat (LTR) retrotransposons (Figure 1.3B) and non-LTR retrotransposons (Figures 1.3C-D).

LTR retrotransposons: LTR retrotransposons are related to retroviruses in structure and function. Human LTR retrotransposons (also known as human endogenous retroviruses or HERVs) (Figure 1.3B) comprise ~8% of the human genome (Lander et al. 2001). Similar to retroviruses, LTR retrotransposons are flanked by a set of LTR sequences, which are direct repeat sequences that encode regulatory elements important for LTR retrotransposon function (Craig et al. 2002, Beauregard et al. 2008). LTR retrotransposons typically encode reverse transcriptase and structural genes analogous to retroviral *pol* and *gag* genes respectively. The LTR retrotransposition mechanism is also similar to retrovirus

replication; however, LTR retrotransposons lack a functional envelope (*env*) gene, which relegates them to an intracellular existence.

Human endogenous retroviruses are not thought to be currently active in humans (*i.e.*, not retrotransposition-competent) (Lander et al. 2001). With that being stated, certain evidence suggests that elevated HERV RNA and protein expression may be associated with certain human diseases including cancer and diabetes (Conrad et al. 1997, Galli et al. 2005, Bannert and Kurth 2006). Although evidence of a retrotransposition competent HERVs has not yet been reported, the existence of human and chimpanzee specific endogenous retroviral insertions suggests the possibility that endogenous retroviruses have been active since the divergence of the human and chimpanzee lineages (Medstrand and Mager 1998, Yohn et al. 2005).

Non-LTR retrotransposons: As their name suggests, non-LTR retrotransposons do not encode LTR sequences. Non-LTR retrotransposons comprise about 35% of human DNA and are the only active group of transposable elements in the human genome (Lander et al. 2001). Examples of human non-LTR retrotransposons include long interspersed nuclear element-1 (LINE-1 or L1) (Figure 1.3C) and Alu elements (also known as short interspersed nuclear elements (SINEs)) (Figure 1.3D). In general, non-LTR retrotransposons may encode one or two open reading frames and end in poly-adenosine rich sequences; thus, they tend to resemble cellular mRNAs in sequence arrangement (Richardson et al. 2015). Non-LTR retrotransposons also typically encode a RT enzyme that is similar to LTR retrotransposon and retroviral RT; however, non-LTR retrotransposon RT is often augmented by an additional endonuclease (EN) enzymatic activity which is important for the non-LTR retrotransposition mechanism (Xiong and Eickbush 1990, Feng et al. 1996, Yang et al. 1999).

Human L1 Elements

L1 is the only active autonomous transposable element in humans. Approximately 516,000 L1 sequences litter the human genome accounting for ~17% of human DNA. Most L1 sequences are incapable of retrotransposition as they are either 5'-truncated, internally rearranged, or extensively mutated (Lander et al. 2001). It is currently estimated, however, that each diploid human genome contains 80 to 100 intact L1 sequences that are capable of retrotransposition (Sassaman et al. 1997, Brouha et al. 2003).

Overview of L1 Structure

A full-length active human L1 is approximately 6 kb in length and consists of a 5' UTR that promotes L1 transcription (Swergold 1990), two open reading frames (ORFs) that are separated by a 63 bp intergenic spacer, and a short 3' UTR that ends in a variable length poly-adenosine tract (Scott et al. 1987, Dombroski et al. 1991) (Figure 1.3C). L1 ORF1 encodes an ~40 kDa nucleic acid binding protein (ORF1p) (Martin 1991, Martin and Branciforte 1993, Hohjoh and Singer 1996, Hohjoh and Singer 1997) that also exhibits nucleic acid chaperone activity (Martin and Bushman 2001, Khazina and Weichenrieder 2009, Khazina et al. 2011, Callahan et al. 2012). L1 ORF2 encodes ORF2p, which contains endonuclease (EN) (Feng et al. 1996, Cost et al. 2001) and reverse transcriptase (RT) activities (Mathias et al. 1991, Dombroski et al. 1994). Both L1 proteins are required for L1 retrotransposition (Moran et al. 1996).

L1 5' UTR

The L1 5' UTR contains internal RNA polymerase II promoter activity that directs transcription of the L1 in the 5' to 3' direction at or near the first L1 nucleotide (Swergold 1990, Becker et al. 1993, Athanikar et al. 2004). The L1 5' UTR also exhibits antisense (3' to 5' with respect to the L1) promoter activity, the function of which has not yet been defined, but in principle could direct the transcription of genes adjacent to the L1 5' UTR (Speek 2001, Macia et al. 2011). Within the 5' UTR are several transcription factor-binding sites including one Yin Yang 1 (YY1)-binding site (Becker et al. 1993, Athanikar et al. 2004), two SRY-

related (SOX)-binding sites (Tchenio et al. 2000), and a Runx3-binding site (Yang et al. 2003). Experiments in cultured cells have shown that deletions or mutations of these binding sites reduce L1 transcription and/or L1 retrotransposition efficiency (Becker et al. 1993, Tchenio et al. 2000, Yang et al. 2003, Athanikar et al. 2004). Notably, mutations in the YY1-binding site results in a loss of L1 transcription fidelity suggesting that YY1 directs transcription to begin at the L1 +1 nucleotide (Athanikar et al. 2004).

L1 ORF1

The first L1 ORF encodes ORF1p, an ~40 kDa nucleic acid binding protein (Martin 1991, Holmes et al. 1992, Hohjoh and Singer 1996, Hohjoh and Singer 1997) with demonstrated nucleic acid chaperone activity (Martin and Bushman 2001). Three ORF1p domains have been identified that are important for ORF1p function. The amino terminal ORF1p domain consists of a coiled-coil domain marked by a putative leucine zipper motif (Holmes et al. 1992, Hohjoh and Singer 1996). Evidence suggests that the coiled-coil domain mediates the formation of an ORF1p trimer, which is the basic ORF1p functional unit (Martin et al. 2003, Basame et al. 2006, Khazina and Weichenrieder 2009). Numerous studies have demonstrated that mutations in the coiled-coil domain are incompatible with retrotransposition, and impair the ability of ORF1p to form proper trimers and to localize properly to L1 RNPs (Moran et al. 1996, Basame et al. 2006, Khazina and Weichenrieder 2009, Doucet et al. 2010, Khazina et al. 2011)

Early studies using mouse and human cultured cells showed that ORF1p binds to L1 RNA to form ribonucleoprotein particles (RNPs) (Martin 1991, Martin and Branciforte 1993, Hohjoh and Singer 1996, Hohjoh and Singer 1997). Genetic and biochemical analysis has demonstrated that the central domain of ORF1p, consists of an RNA recognition motif (RRM) that interacts with the ORF1p carboxyl terminal domain (CTD) to mediate nucleic acid binding (Januszyk et al. 2007, Khazina and Weichenrieder 2009, Khazina et al. 2011). Critically, mutation of either the RRM or CTD disrupts ORF1p nucleic acid binding activity, prevents proper L1 RNP formation, and renders L1 unable to

retrotranspose (Moran et al. 1996, Martin et al. 2000, Kulpa and Moran 2005, Martin et al. 2005, Goodier et al. 2007, Khazina and Weichenrieder 2009, Doucet et al. 2010, Khazina et al. 2011).

In addition to nucleic acid binding, ORF1p exhibits nucleic acid chaperone activity that facilitates the dissociation of mismatched DNA duplexes and/or the re-annealing of complimentary DNA strands *in vitro* (Martin and Bushman 2001, Martin et al. 2005, Khazina and Weichenrieder 2009, Callahan et al. 2012). Certain genetic and biochemical data suggest that ORF1p nucleic acid chaperone activity is required for L1 retrotransposition; specifically that ORF1p nucleic acid chaperone activity may be important for the integration of L1 into genomic DNA (Martin and Bushman 2001, Martin et al. 2005, Martin et al. 2008). Data from the study of other non-LTR retrotransposons, however, suggests that ORF1p is not needed for retrotransposition. For example, the zebrafish LINE (Zfl2-1), which like human L1 also encodes an ORF1p protein with chaperone activity, does not strictly require ORF1p for retrotransposition (Kajikawa et al. 2012, Nakamura et al. 2012). Additionally, other non-LTR retrotransposons such as Alu elements do not strictly require ORF1p to retrotranspose (Dewannieux et al. 2003), although it has been proposed the ORF1p may enhance Alu activity (Wallace et al. 2008). It is evident that the development of *in vitro* systems that are able to test specific steps in the retrotransposition pathway (*i.e.*, TPRT integration step) would provide a better test for what ORF1p functions are necessary for retrotransposition.

L1 ORF2

The second L1 ORF encodes an ~150 kDa protein, ORF2p (Ergun et al., 2004; Doucet et al., 2010; Goodier et al., 2010). ORF2p exhibits endonuclease (EN) (Feng et al., 1996) and reverse transcriptase (RT) (Dombroski et al., 1994; Mathias et al., 1991) activities, which are required for L1 retrotransposition (Moran et al. 1996). The amino terminal portion of ORF2p encodes apurinic/apyrimidinic (AP) endonuclease domain (Feng et al. 1996). ORF2p EN preferentially cleaves single stranded DNA at an AT-rich consensus sequence: 5'

TTTT/A 3' where the "/" indicates the cleavage site (Feng et al. 1996). Mutations to catalytic residues within the EN domain impair the ability for ORF2p to cleave DNA and renders L1 unable to retrotranspose.

ORF2p RT activity resides within a domain located immediately downstream of the EN domain. The ORF2p RT domain is similar to RT domains from group II introns, retroviruses, telomerase and other endogenous retrotransposons (Hattori et al. 1986, Xiong and Eickbush 1990, Eickbush 1997, Malik et al. 1999). Early experiments using ORF2 fusion proteins expressed in yeast first demonstrated that ORF2 exhibited RT activity (Mathias et al. 1991, Dombroski et al. 1994). Subsequent studies were able to purify recombinant ORF2p from insect cells and to demonstrate that recombinant ORF2p exhibited RNA-dependent and DNA-dependent DNA polymerase activities (Piskareva et al. 2003). Notably, ORF2 RT is highly processive and lacks a detectable RNase H activity; traits which distinguish it from retroviral RTs (Piskareva et al. 2003, Piskareva and Schmatchenko 2006). Later studies went on to show that ORF2p RT activity co-purified with L1 RNPs and that L1 RNP-associated ORF2p RT activity could use the L1 RNA, and to a lesser extent other cellular RNAs, as a template for reverse transcription (Kulpa and Moran 2006, Doucet et al. 2010).

The carboxyl portion of ORF2p is rich in cysteine amino acid residues and is thought to encode a potential zinc-knuckle structure (Fanning and Singer 1987). The function of the cysteine-rich (C-rich) domain is unknown; however, mutations in this region negatively impact L1 retrotransposition, suggesting that it is important for ORF2p function (Moran et al. 1996). Notably, mutations in the C-rich domain also negatively impact ORF2 RT activity and localization of ORF2p to L1 RNPs (Clements and Singer 1998, Doucet et al. 2010). Evidence from a recent study using a purified peptide from the carboxyl portion of ORF2p suggests that the ORF2p carboxyl-terminus may be involved in RNA binding (Piskareva et al. 2013). Clearly, the development of *in vitro* assays to test C-rich domain function and/or structural analyses similar to those characterizing the ORF2p EN domain (Weichenrieder et al. 2004, Repanas et al. 2007) will help further elucidate the function of the C-rich domain in retrotransposition.

L1 3' UTR

The L1 3' UTR is ~200 bp length and is located between the stop codon of ORF2 and the variable length poly adenosine region at the 3' end of the element (Grimaldi et al. 1984, Scott et al. 1987, Dombroski et al. 1991). The 3' UTR varies significantly between L1s from different mammals except for a highly conserved G-rich polypurine tract that is predicted to form a stable secondary structure (Howell and Usdin 1997). The L1 3' UTR also contains a relatively weak polyadenylation signal that is often bypassed in favor of a stronger polyadenylation signal in flanking DNA (Holmes et al. 1994, Moran et al. 1996, Moran et al. 1999). Notably, the L1 3' UTR is not strictly required for retrotransposition as cell culture experiments have demonstrated that the native L1 polyadenylation signal can be efficiently bypassed and that mutant L1 elements with deletions in the 3' UTR encompassing the G-rich tract retrotranspose at efficiencies similar to wild-type L1 elements (Moran et al. 1996, Moran et al. 1999).

The L1 Retrotransposition Pathway

Overview

The L1 retrotransposition cycle (Figure 1.4) involves the transcription of a genomic L1 element, which produces a bicistronic L1 mRNA that is then translated in the cytoplasm by an unconventional cap-dependent mechanism (Leibold et al. 1990, McMillan and Singer 1993, Alisch et al. 2006, Dmitriev et al. 2007). Following translation, ORF1p and ORF2p preferentially bind to their encoding L1 mRNA (a phenomenon termed *cis*-preference (Esnault et al., 2000; Wei et al., 2001)) to form a L1 ribonucleoprotein particle (RNP), which is an important L1 retrotransposition intermediate (Doucet et al., 2010; Hohjoh and Singer, 1996; Kulpa and Moran, 2005, 2006; Martin, 1991). The L1 RNP is thought to gain entry to the nucleus by a process that does not strictly require cell division (Kubo et al., 2006), although L1 retrotransposition seems to be enhanced in dividing cells (Shi et al., 2007; Xie et al., 2013). Once components

of the L1 RNP have entered the nucleus, the L1 RNA is reverse transcribed and is integrated into genomic DNA by a concerted process termed target-site primed reverse transcription (TPRT) (Cost et al., 2002; Feng et al., 1996; Luan et al., 1993). Some of these steps will be discussed in further detail below.

Trans-mobilization of non-autonomous elements

The L1-encoded proteins can work in *trans* to promote the retrotransposition of non-autonomous non-LTR retrotransposons such as Alu elements (Figure 1.3D) (Dewannieux et al. 2003) and SINE-R/VNTR/Alu (SVA) elements (Ostertag et al. 2003, Hancks et al. 2011, Raiz et al. 2012). Alu elements are ~300 bp sequences derived from 7SL RNA and are present in over one million copies in the genome (~11% of human DNA) (Deininger et al. 1981, Ullu and Tschudi 1984, Sinnett et al. 1991, Chu et al. 1995, Lander et al. 2001). SVA (SINE-R/VNTR/Alu) elements are composite sequences derived from retrotransposon (SINE-R), variable length tandem repeat (VNTR), and Alu sequences, and are present in ~2,700 copies in the human genome (Ostertag et al. 2003, Wang et al. 2005). Alu and SVA elements are considered non-autonomous because they do not encode the proteins necessary to promote their own retrotransposition and must therefore parasitize the L1-encoded proteins in order to retrotranspose.

In addition to Alu and SVA elements, a number of other cellular RNAs also can be retrotransposed by the L1-encoded proteins. These RNA sequences include uracil-rich small nuclear RNAs (e.g., U6 snRNA (Buzdin et al. 2002, Gilbert et al. 2005, Garcia-Perez et al. 2007)), small nucleolar RNAs (e.g., U3 snoRNA (Weber 2006)), and messenger RNAs (Figure 1.3E) (Esnault et al. 2000, Wei et al. 2001). Taking into account all known RNA species that are mobilized *in trans* by the L1 machinery, the process of L1 retrotransposition has generated nearly 1 billion base pairs of human DNA (Lander et al. 2001).

L1 Translation

The bicistronic nature of the L1 mRNA sets it apart from other eukaryotic mRNAs, which typically encode a single ORF, and leads to the question of how

the L1 RNA is efficiently translated. Early studies demonstrated that L1 ORF1p was expressed in certain cancer derived human and mouse cell lines and/or cells transfected with L1 cDNA, which provided the initial evidence that L1 RNA was translated (Leibold et al. 1990, Holmes et al. 1992, McMillan and Singer 1993). In contrast to the relative ease in detecting ORF1p, L1 ORF2p was much more difficult to detect. The only evidence that supported the translation of ORF2 was provided by 1) indirect biochemical evidence derived from in vitro assays that measured ORF2p RT activity or 2) from genetic cell culture-based retrotransposition assays (Mathias et al. 1991, McMillan and Singer 1993, Moran et al. 1996).

Despite the inability to directly detect ORF2p, genetic and biochemical analyses demonstrated that human L1 mRNA was likely translated in a cap-dependent manner that involved an unconventional termination/re-initiation mechanism (Alisch et al. 2006, Dmitriev et al. 2007). This model of human L1 translation postulates that after reaching the termination codon of ORF1, a ribosome continues to scan beyond the end of ORF1 and then re-initiates translation at the first ORF2 methionine (AUG) codon (Alisch et al. 2006). Another model of L1 translation based on studies of mouse L1 elements proposes that mouse L1 translation is dependent on an internal ribosome entry site (IRES) located upstream of both ORFs (Li et al. 2006). Studies of human L1 translation, however, suggest this is likely not the case for human L1 as the entire ORF1 sequence can be deleted or replaced by green fluorescence protein (GFP) without affecting ORF2 translation (Alisch et al. 2006). Recent development of epitope tagging strategies (Goodier et al. 2004, Doucet et al. 2010) to aid in the purification and detection of L1 proteins should prove useful in future studies aimed at determining how L1 is translated.

The L1 RNP

The L1 RNP minimally is composed of L1 RNA, ORF1p, and ORF2p and is a necessary L1 retrotransposition intermediate. Early studies first suggested that L1 RNA and ORF1p formed an RNP by demonstrating that L1 ORF1p and

L1 RNA expressed in mouse F9 cells (embryonal carcinoma cells) co-sediment together in sucrose gradients (Martin 1991, Martin and Branciforte 1993). Subsequent studies using cultured human cell lines showed that human ORF1p binds to human L1 RNA to form RNPs (Hohjoh and Singer 1996, Hohjoh and Singer 1997, Hohjoh and Singer 1997) and that L1 RNP formation was necessary for L1 retrotransposition (Kulpa and Moran 2005).

It remained unknown for some time, however, whether L1 ORF2p was also part of the L1 RNP due to an inability to detect ORF2p using standard biochemical methods (Goodier et al. 2004). Eventually, an *in vitro* assay designed to detect L1 ORF2 RT activity (L1 element amplification protocol or LEAP) (Kulpa and Moran 2006) was used to show that ORF2p RT activity co-purified with L1 RNA and ORF1p in cells transfected with engineered L1 expression plasmids. Subsequent studies using epitope tagged engineered L1 plasmids were able to demonstrate a direct physical interactions between ORF2p and L1 RNA and ORF1p, thus augmenting previous studies that used enzymatic assays to infer the presence of ORF2p in L1 RNPs (Doucet et al. 2010, Goodier et al. 2010). Despite these findings, the stoichiometry of L1 RNA and L1-encoded proteins within an L1 RNP and whether or not other host cell proteins are required for L1 RNP function remains unknown. Notably, a recent study has demonstrated that different types of L1 RNPs may exist, suggesting that the composition of L1 RNPs may change during the L1 retrotransposition cycle (Taylor et al. 2013). Indeed, the development of an *in vitro* cell-free system that uses purified L1 RNA and proteins could be useful in determining the composition and biochemical properties of functional L1 RNPs.

TPRT

The reverse transcription of L1 RNA and subsequent integration of L1 into genomic DNA occurs at the genomic DNA target site by a process termed target site primed reverse transcription (TPRT), which was initially characterized for the R2Bm non-LTR retrotransposon from the silkworm moth (Luan et al. 1993). In the case of L1, once components of the L1 RNP enter the nucleus, the ORF2 EN

is thought to generate a single-strand endonucleolytic nick in a single strand of genomic DNA at a thymidine rich consensus sequence (e.g., 5'-TTTT/A, 5'-TCTT/A, 5'-TTTA/A, etc. where the '/' indicates the cut site) (Cost and Boeke, 1998; Feng et al., 1996; Morrish et al., 2002). This cleavage event exposes a free 3' hydroxyl group, which can then be used by the ORF2p RT as a primer to initiate antisense (-) strand L1 cDNA synthesis using the L1 mRNA as a template for the reverse transcription reaction (Cost and Boeke, 1998; Feng et al., 1996). Following L1 (-) cDNA synthesis, the top strand of DNA must then be cleaved and the L1 cDNA must be copied to complete integration. How these last steps of TPRT occur are not fully understood, but it is thought these steps may involve ORF2 EN/RT and/or host proteins involved in DNA repair (Morrish et al., 2002; Suzuki et al., 2009; Taylor et al., 2013). The TPRT process typically results in the generation of variable length target site duplications that flank new L1 insertions, but can also result in other target site alterations including target site deletions (Gilbert et al. 2002, Symer et al. 2002, Gilbert et al. 2005, Richardson et al. 2015).

The Impact Of L1 Retrotransposition On The Human Genome

Discovery of an active human L1

It was the discovery that L1 retrotransposition caused hemophilia A that alerted researches that L1 was active in the human genome (Kazazian et al. 1988). In 1988, researchers from the Kazazian lab conducted a screen to detect abnormalities in the coagulation factor VIII (*F8*) gene of a cohort of 240 male patients stricken with hemophilia A. The screen revealed L1 insertions into exon 14 of the *F8* gene in 2 out of the 240 patients. Biochemical testing revealed that the *F8* gene was not functional in either of the affected patients suggesting that the L1 insertions had inactivated the *F8* gene. Critically, genetic testing revealed that the *F8* gene was normal (*i.e.*, did not contain an L1 insertion) in the parents of the affected individuals. This suggested that the L1 had retrotransposed into the *F8* genes of the affected patients either the parental germline or during

development (Kazazian et al. 1988). In two subsequent studies, a full-length L1 element (L1.2) (Dombroski et al. 1991) that likely produced one of the disease-causing *F8* gene insertions was cloned and then demonstrated to actively retrotranspose in cultured human cells (Moran et al. 1996). Together, these landmark studies confirmed that L1 was active in humans and that L1 retrotransposition could have a profound impact on the human genome.

L1 and human disease

L1 retrotransposition is responsible for at least 96 single-gene genetic diseases in humans (Hancks and Kazazian 2012). Insertional mutagenesis caused by the insertion of L1, Alu or SVA (Figures 1.5A and 1.5D) sequences into genes can cause mutations (*e.g.*, nonsense mutations, frame-shift mutations) that can disrupt normal gene expression and lead to disease. L1-mediated insertional mutagenesis has been documented in numerous human disease cases including hemophilia A (Kazazian et al. 1988), Duchene muscular dystrophy (Yajima et al. 1999), and cancer (Miki et al. 1992). Indeed, the suppression of gene expression by transposable element insertions is what lead to the initial discovery of transposable elements by Barbara McClintock (McClintock 1950, McClintock 1984) and the discovery that L1 was active in the human genome (Kazazian et al. 1988). Notably, L1-mediated insertional mutagenesis is also responsible for disease phenotypes in other animals. For example, mutations resulting from L1 retrotransposition are responsible for several mutant mouse strains including the spastic mouse (Mulhardt et al. 1994), *med* (motor endplate disease) mouse (Burgess et al. 1995, Kohrman et al. 1996) and black-eyed white mouse (Yajima et al. 1999). L1 retrotransposition is also associated with diseases in dogs including hemophilia B and narcolepsy (Lin et al. 1999, Brooks et al. 2003).

Genomic DNA deletions: L1 retrotransposition can sometimes result in alterations of genomic DNA at the L1 insertion site (Gilbert et al. 2002, Symer et al. 2002, Gilbert et al. 2005, Han et al. 2008). For example, studies using human cultured cells have demonstrated that L1 retrotransposition can sometimes result

in genomic target-site deletions ranging from 193 bp to ~3.1 kb (Figure 1.5B) (Gilbert et al. 2002, Gilbert et al. 2005). Notably, L1-mediated deletions are associated with cases of human disease. For example, an L1 retrotransposition event was associated with a large interstitial deletion of ~46 kb in the *PDHX* gene of a patient diagnosed with pyruvate dehydrogenase complex deficiency (Mine et al. 2007). To date, L1 and Alu retrotransposition mediated deletions are associated with 12/96 (~13%) of L1 retrotransposition-associated human disease cases (Chen et al. 2005, Hancks and Kazazian 2012).

Modulation of gene expression: L1 elements encode a number of active splice acceptor/donor sites as well as several polyadenylation signals that could cause the missplicing and/or premature polyadenylation of cellular genes that contain L1 insertions (Figure 1.5D) (Perepelitsa-Belancio and Deininger 2003, Han et al. 2004, Belancio et al. 2006). Analysis of several human disease cases has revealed that L1 insertions may interfere with the normal processing of cellular RNAs. For example, in a patient diagnosed with choroideremia, an L1 insertion into the choroideremia (*CHM*) gene was associated with an exon-skipping event that deleted exon 6 from the *CHM* mRNA (van den Hurk et al. 2003). Alu and SVA insertions also have been associated with diseases involving defects in normal RNA processing. For example, an Alu insertion was associated with the skipping of exon 22 of the *BRCA2* tumor suppressor gene in a patient with breast cancer (Miki et al. 1996) and an SVA insertion into the 3' UTR of the *fukutin* gene was associated with the truncation of the *fukutin* mRNA in a patient with Fukuyama muscular dystrophy (FCMD) (Taniguchi-Ikeda et al. 2011).

Post-insertion genomic rearrangements: Due to the widely dispersed number of L1 and Alu repeats throughout the genome, unequal crossing over (*i.e.*, non-allelic homologous recombination (NAHR)) between these sequences can sometimes occur, which can result in chromosomal DNA deletions, duplications and/or inversions (Figure 1.5F) (Fitch et al. 1991, Gilbert et al. 2002, Gilbert et al. 2005, Han et al. 2008, Lee et al. 2008). NHAR between L1 and Alu repeats have been documented in several human disease cases including Alport syndrome, Perlman syndrome, and a case of glycogen storage disease that was

associated with a 7574 bp deletion within the patient's phosphorylase kinase beta (*PHKB)* gene caused by an unequal cross-over event between non-allelic L1s (Burwinkel and Kilimann 1998, Segal et al. 1999, Hancks and Kazazian 2012, Higashimoto et al. 2013).

L1 and human genome evolution

As discussed up to this point, L1 retrotransposition can cause mutations that cause diseases; clearly a negative impact on the human genome. If however, a new L1 insertion does not result in a life-threatening disease, the insertion may become a permanent feature of the human genome. This is evident from the large number of L1 and Alu sequences extant in the genome. What then, if any, is the significance of the increasing accumulation of new L1 retrotransposition-mediated sequence insertions (*e.g.*, L1, Alu, SVA, *etc.*) on the human genome?

Active L1 retrotransposition in human genomes is a mechanism that generates genomic sequence diversity in the human population. Indeed, it is currently estimated that a new L1 retrotransposition event occurs once in every 20 to 200 live human births while the current Alu retrotransposition rate is estimated to occur 1 out of every 20 births (Cordaux and Batzer 2009). Recent comparative analyses between the human genome reference sequence (HGR) and individual genomes representing distinct human populations have revealed the existence of rare active L1 alleles in the human population that are not present in the HGR, which suggests that the number and diversity of active L1s present in the human population are likely underrepresented by the HGR (Beck et al. 2010, Ewing and Kazazian 2010). Notably, recent studies have uncovered compelling new evidence that L1 retrotransposition also occurs in adult somatic tissues including the human brain, which suggests that L1 retrotransposition contributes to somatic variation within an individual (Coufal et al. 2009, Baillie et al. 2011, Upton et al. 2015). Thus, L1 retrotransposition generates intra- and inter-individual diversity in the human population.

L1-mediated insertions into coding regions could introduce novel regulatory elements that could potentially modulate normal gene expression (Cordaux and Batzer 2009). Analysis of transposable elements sequences in the human genome has revealed that approximately 4% of protein coding genes may contain transposable element insertions with L1 and Alu insertions accounting for nearly 70% of those insertions (Gibbs et al. 2004). L1 insertions into gene coding regions could provide novel transcription factor binding sites, splice sites and/or polyadenylation signals, which could directly affect gene expression. Experiments done in cultured human cells suggest that L1 insertions within protein coding genes can attenuate gene expression, which may involve premature polyadenylation and/or other transcriptional elongation defects (Perepelitsa-Belancio and Deininger 2003, Han et al. 2004). Additional studies suggest that L1 and Alu sequences could introduce alternative splice sites into coding regions that may affect the way cellular mRNAs are processed (Belancio et al. 2006, Belancio et al. 2008). Finally, recent large-scale sequence analyses of the human and mouse transcriptomes have revealed that retrotransposon sequence insertions at the 5' end of protein-coding genes can function as alternate promoters and that retrotransposon insertions into the 3' UTRs of protein coding genes are associated with decreases in gene expression (Faulkner et al. 2009).

L1 retrotransposition and genomic structural variation: Illegitimate recombination events between L1 and Alu repeats have made a significant impact on human genome structure. Combined analyses of the human genome suggest that ectopic recombination between Alu sequences may have played a significant role in the expansion of segmental duplications in the human genome, which account for ~4-5% of genomic DNA (Fitch et al. 1991, Bailey et al. 2003, Han et al. 2008, Lee et al. 2008). A potential significant biological outcome of segmental duplications is that these events could result in the duplication of cellular genes (Bailey et al. 2002, Long et al. 2003). Notably, ectopic recombination between repeat sequences are postulated to have given rise to human fetal globin genes (Shen et al. 1981). Additionally, unequal crossing over

between L1 sequences is thought to be responsible for the duplication of the γ-globin gene in the primate lineage (Fitch et al. 1991), and may have been responsible for the *Abp* gene duplication in the mouse (Janousek et al. 2013).

L1 and Alu-mediated deletions and inversions also have changed the genomic DNA landscape. Recent comparative genomic analyses between the human and chimpanzee have uncovered at least 73 L1 recombination-associated deletions and 492 Alu recombination-associated deletions that are unique to the human genome (Han et al. 2005, Sen et al. 2006, Han et al. 2008). Additional sequence comparisons between the human and chimpanzee genomes have revealed that L1 and Alu sequences are associated with 112 out of 252 identified chromosomal inversions between humans and chimpanzee (Lee et al. 2008). Notably, 27 L1 and/or Alu associated inversions are human-specific and 22 are chimpanzee-specific (Lee et al. 2008). Thus, illegitimate recombination events between L1 and Alu are responsible for genomic structural changes that may have contributed to differences between the human and chimpanzee genomes.

Retrotransposition of other cellular RNAs: L1 retrotransposition generates additional genomic sequence diversity through the retrotransposition of other cellular RNAs such as mRNA (Esnault et al. 2000, Wei et al. 2001) and non-coding RNAs (*e.g.*, U6 snRNA, U3 snoRNA, Y RNA) (Buzdin et al. 2002, Perreault et al. 2005, Weber 2006). The retrotransposition of mRNA, for example, can give rise to processed pseudogenes, which are typically defined as non-functional intronless copies of functional genes. Comprehensive sequencing analysis has revealed that the human genome contains approximately 8,000 processed pseudogenes, which originate from about 2,500 different functional genes (Zhang et al. 2003, Pei et al. 2012). Even though processed pseudogenes by definition are not functional, some processed pseudogenes are expressed perhaps through the recruitment of adjacent promoters or as part of other transcripts (Strichman-Almashanu et al. 2003, Vinckenbosch et al. 2006, Pink et al. 2011, Pei et al. 2012).

Remarkably, the retrotransposition of cellular RNAs may have given rise to functional genes. For example, the human phosphoglycerate kinase 2 (*PGK2*) gene is a testis specific intronless gene on chromosome 6 that likely resulted from the retrotransposition of an mRNA that originated from the *PGK1* gene located on the X chromosome (McCarrey and Thomas 1987, McCarrey 1990). Notably, a number of retrotransposed mammalian testis-specific genes have been identified (Bradley et al. 2004, Marques et al. 2005). A recent analysis of the canine genome has revealed that a retrotransposed copy of the fibroblast growth factor 4 gene (*fgf4*) is associated with chondrodysplasia (also known as short-limbed or disproportional dwarfism) in over 20 domesticated dog breeds (Parker et al. 2009).

L1 transductions: A L1 transduction involves the retrotransposition of genomic DNA sequences from either the 5' and 3' ends of genomic L1 (Figure 1.5C). The retrotransposition of genomic DNA from the L1 3' end (known as an L1 3' transduction) can occur if L1 transcription continues past the 3' end of an L1 into the adjoining genomic DNA. This produces a chimeric RNA transcript consisting of the L1 element and genomic DNA sequences from the 3' end of the L1. The subsequent retrotransposition of this chimeric L1/3' genomic DNA transcript would result in the transduction of the 3' flanking genomic DNA sequence to a new location in the genome (Holmes et al. 1994, Moran et al. 1996).

L1-mediated 3' transductions have been detected in human disease cases (Miki et al. 1992, Holmes et al. 1994). For example, a L1 3' transduction was discovered in the dystrophin (*DMD*) gene of a patient that was diagnosed with Duchene muscular dystrophy (Holmes et al. 1994). In this case, the L1 3' transduction consisted of 1,400 bp of L1 sequence, followed by a 37 bp poly-adenosine tract, and an additional 489 bp of non-L1 sequence that was followed by another poly-adenosine tract. The unique 489 bp sequence was mapped back to an identical genomic DNA sequence on chromosome 1, which resided immediately downstream of a full-length genomic L1 element (LRE2). Thus, the 3' transduced DNA sequence facilitated the identification of the active progenitor

element (LRE2) that produced the *DMD* disease insertion. The subsequent use of cell culture based L1 retrotransposition assays demonstrated that L1 retrotransposition could give rise to 3' transductions in cultured human cells (Moran et al. 1996). Notably, recent analyses of the human genome sequence have revealed that L1-mediated 3' transductions are responsible for approximately 0.6-1.0% of the human genome (Goodier et al. 2000, Pickeral et al. 2000) and the use of L1-mediated 3' transduction genomic DNA "tags" have enabled researchers to identify highly active L1 subfamilies and explore the dynamics of L1 lineage evolution (Brouha et al. 2002, van den Hurk et al. 2007, Beck et al. 2010, Macfarlane et al. 2013).

In addition to 3' transductions, genomic DNA at the 5' end of an L1 can be retrotransposed (known as an L1 5' transduction) if for example, L1 transcription were to initiate upstream of the L1 5' end. Indeed, analysis of the human genome and direct sequencing of L1 insertions derived from engineered human L1 retrotransposition insertions in cultured cells have confirmed that L1 5' transductions can occur *in vivo* and in culture cells (Lander et al. 2001, Wei et al. 2001, Symer et al. 2002). Notably, only two L1 5' transductions have been identified in the HGR, suggesting that L1 5' transductions do not occur at the same frequency as L1 3' transductions. This could be due to that fact that the majority of genomic L1 insertions are 5' truncated. Notably, a somatic L1 5' transduction was recently detected in the human brain (Evrony et al. 2012), and approximately 8% of SVA elements in the human genome are associated with 5' transductions (Damert et al. 2009, Hancks et al. 2009, Hancks et al. 2011). L1 5' transductions also have been detected in other animals. For example, an L1-mediated 5' transduction is associated with an inactivating mutation in the *NR2E3* gene in the retinal degeneration 7 (rd7) mouse, (Symer et al. 2002, Chen et al. 2006).

L1-mediated exon shuffling: The concept of exon shuffling arose from the recognition that genes were split into coding (*i.e.*, exons) and non-coding (*i.e.*, introns) regions and suggests that exons from different genes can be brought together by recombination events to generate sequences that encode novel

proteins (Sambrook 1977, Gilbert 1978). Experiments carried out in cultured human cells first demonstrated that L1-mediated transduction could in principle mediate exon shuffling. In these experiments, researchers demonstrated that a synthetic exon could be transduced into a cellular gene via L1-mediated 3' transduction and that the synthetic exon could then be expressed as a fusion product with that gene (Moran et al. 1999). Notably, three *AMAC* gene duplications are thought to have resulted from SVA-mediated 3' transductions (Rozmahel et al. 1997, Xing et al. 2006) and recent high-throughput sequencing of human cancer genomes has revealed that nearly one quarter of cancer associated L1 insertions contain 3' transductions, which occasionally involve entire exons and/or gene regulatory elements (Tubio et al. 2014).

The evolution of new genes via retrotransposition: Recent studies also have revealed that L1 retrotransposition may have played a role in the evolution of genes involved in HIV-1 resistance in distantly related primate species. In owl monkeys for example, a TRIM5-cyclophillin A fusion protein (TRIMCyp) that confers resistance to HIV-I was determined to have resulted from the retrotransposition of cyclophilin A (*CypA*) mRNA into exon 7 of the *TRIM5* gene (Nisole et al. 2004, Sayah et al. 2004). In macaques, a similar TRIMCyp fusion protein was also discovered. Unlike the owl monkey TRIMCyp fusion, the macaque TRIMCyp fusion resulted from retrotransposition of *CypA* mRNA into the 3' UTR of the *TRIM5* gene (Brennan et al. 2008, Virgen et al. 2008, Wilson et al. 2008). Thus, the L1 retrotransposition may have contributed to convergent evolutionary processes in separate branches of the primate lineage.

Where And When Does L1 Jump?

L1 must retrotranspose in the germline or in cells that contribute directly to the germline to ensure its continued existence. Indeed many studies have provided evidence that L1 retrotransposition does occur in the germline as well as in somatic cells during embryogenesis. Recent studies now suggest that L1

retrotransposition also occurs in certain adult somatic tissues including cells of the mammalian nervous system.

Germline retrotransposition

Early studies provided evidence suggesting that L1 retrotransposition occurs in the germline by demonstrating that L1 proteins and L1 RNA are expressed in cell lines derived from mouse reproductive tissues (Branciforte and Martin 1994, Trelogan and Martin 1995, Soper et al. 2008). Subsequent studies showed that L1 proteins are expressed in the reproductive tissues of human males (Ergun et al. 2004) and that L1 RNA is expressed in human oocytes (Georgiou et al. 2009). Cultured human oocytes have also been demonstrated to support the retrotransposition of engineered human L1 elements (Georgiou et al. 2009). The analysis of human diseases further suggests that L1 retrotransposition may take place in the human germline. For example, a case of chronic granulomatous disease (CGD) was thought to have been caused by a mutagenic L1 insertion into the *CYBB* gene during maternal meiosis I or possibly during early development (Brouha et al. 2002).

Due to the fact that it is difficult to track L1 retrotransposition events in live humans, transgenic animal models have been developed to facilitate the study of L1 retrotransposition *in vivo.* (Ostertag et al. 2002, An et al. 2006, Kano et al. 2009, O'Donnell et al. 2013). L1-transgenic animals carry an engineered L1 transgene that is typically marked with retrotransposition indicator cassette that expresses a fluorescent marker protein (*i.e.*, EGFP), which conveniently allows retrotransposition events to be tracked using EGFP fluorescence or by biochemical methods (*e.g.*, Southern blotting, PCR, *etc.*) using oligonucleotide probes specific to the retrotransposition indicator cassette (Ostertag et al. 2002). Indeed, experiments using L1 transgenic mice have been used demonstrated that engineered L1 transgenes retrotranspose in the germline of male mice (Ostertag et al. 2002).

L1 retrotransposition during development

L1 retrotransposition also can occur in somatic cells during embryonic development. For example, cell culture studies have demonstrated that certain human embryonic stem cell lines (hESC), human and mouse embryonic carcinoma (EC) derived cell lines, as well as mouse post-implantation embryos express endogenous L1 RNA and proteins (Skowronski and Singer 1985, Skowronski et al. 1988, Martin 1991, Trelogan and Martin 1995, Garcia-Perez et al. 2007, Garcia-Perez et al. 2010). Additionally, studies using cell culture based retrotransposition assays have demonstrated that hESCs and human embryonic carcinoma derived cell lines support high levels of L1 retrotransposition from transfected engineered human L1 constructs (Garcia-Perez et al. 2007, Garcia-Perez et al. 2010).

The study of human disease cases has also provided evidence that L1 retrotransposition can occur during development. For example, an embryonic L1 retrotransposition event is thought to be responsible for a case of X-linked choroideremia (van den Hurk et al. 2003, van den Hurk et al. 2007). In this particular case, a male patient diagnosed with X-linked choroideremia carried a mutagenic L1 insertion within the choroideremia (*CHM)* gene that caused aberrant splicing of the *CHM* mRNA (van den Hurk et al. 2003, van den Hurk et al. 2007). PCR analysis of the affected patient's family members revealed that the mother was a somatic and germline mosaic for the mutated L1 *CHM* allele that found in the patient, thus the original mutagenic *CHM* L1 insertion must have occurred in the mother during her development.

Studies using transgenic rodents have also demonstrated that L1 retrotransposes during development (Prak et al. 2003, Muotri et al. 2005, Kano et al. 2009). In a recent study by Kano and colleagues (Kano et al. 2009), a comparison between genomic DNA isolated from L1 transgenic mouse sperm and from pooled pre-implantation embryos revealed that L1 transgene insertions occurred more frequently in pre-implantation embryos. This result suggested that L1 retrotransposition may take place more frequently during embryogenesis than was previously thought. Notably, it was also observed that L1 transgene RNA produced in germline cells could be delivered to the zygote upon fertilization and

subsequently inserted into genomic DNA during embryonic development (Kano et al. 2009). Thus, L1 transcription that takes place in the germline may generate L1 RNA that can be retrotransposed in the developing embryo.

Somatic L1 retrotransposition in the brain

Recent evidence suggests that somatic L1 retrotransposition may influence nervous system development. Some of the initial evidence that L1 retrotransposition occurs in the nervous system was provided by a study that examined L1 expression in the brains of rodents (Muotri et al. 2005). The researchers in this study demonstrated that neural progenitor cells (NPCs) isolated from the rat hippocampus expressed elevated levels endogenous L1 RNA compared to other cells from the same region of the brain (Muotri et al. 2005). Further analyses revealed that rat NPCs were also able to support the retrotransposition of engineered human L1 elements. In addition to data derived from rat NPCs, the researchers in this study demonstrated that engineered L1 transgene retrotransposition events could be detected in the brains of L1 transgenic mice suggesting the somatic L1 retrotransposition occurs in the mammalian brain (Muotri et al. 2005). In a following study that examined L1 retrotransposition in the human brain, researchers showed that human NPCs express elevated levels of endogenous L1 RNA and that human NPCs can support the retrotransposition of engineered human L1 elements (Coufal et al. 2009). Quantitative PCR analysis demonstrated that genomic DNA isolated from the brains of adult human cadavers contained higher L1 copy numbers then genomic DNA isolated from heart and liver tissues (Coufal et al. 2009), suggesting that somatic retrotransposition occurs in the human brain.

A recent study using single neuron sequencing has estimated that there could be as many as 13.7 somatic L1 insertions per hippocampal neuron in the human brain (Upton et al. 2015). This estimate is in contrast to another study that also used single cell sequencing, which provided an estimate of <0.1 insertion per cortical neuron (Evrony et al. 2012) and a study that used quantitative L1 PCR estimated that there could be as many as 80 somatic L1 insertions per brain

cell (Coufal et al. 2009). Despite these differences, these studies help to confirm that L1 is active the human brain; advances in sequencing technologies will no doubt improve the accuracy of future estimates of somatic L1 insertions in the brain. Notably, studies have shown that some somatic L1 insertions in the human brain map to genes with potential neurobiological functions (Baillie et al. 2011) and a recent study has used somatic L1 retrotransposition events to track cell lineages in the frontal cortex providing insight into the temporal and spatial patterns of mosaicism in the human brain (Evrony et al. 2015). In addition to data obtained from studies of the mammalian brain, a recent study in flies has detected elevated transposon activity in the brains of mutant flies, and has identified over 200 *de novo* brain specific transposon insertions, some of which occur into genes involved neuronal functions such as memory (Li et al. 2013, Perrat et al. 2013). These combined data suggest that genomic heterogeneity due to transposon activity may be a conserved feature of the central nervous system. Notwithstanding, the extent of L1 retrotransposition activity and whether L1 retrotransposition affects the development and/or function of the nervous system will need to be carefully addressed in future studies.

L1 and cancer

Evidence also suggests that spontaneous somatic L1 insertions may be a feature of certain types of cancer. For example, early studies identified tumor specific L1 insertions into the *APC* gene of a patient diagnosed with colon cancer (Miki et al. 1992) and into the *myc* gene of a patient diagnosed with breast cancer (Morse et al. 1988). Recently, several labs have used high-throughput sequencing strategies to detect somatic L1 insertions in tumor tissues. Sequencing data from these studies have demonstrated that somatic L1 and Alu insertions are associated with several types of cancers including lung, colon, liver, prostate, ovary, and breast cancers (Iskow et al. 2010, Lee et al. 2012, Solyom et al. 2012, Shukla et al. 2013, Helman et al. 2014, Tubio et al. 2014). Notably, some tumor-specific L1 insertions have been detected within protein coding genes that are commonly mutated in cancer including genes that may function as tumor suppressors (Miki et al. 1992, Lee et al. 2012, Shukla et al.

2013, Helman et al. 2014). Despite these findings, whether L1 mediated insertions are driving tumor formation or merely benign passenger mutations remains to be determined.

The Host Response To L1 Retrotransposition

As discussed in the previous sections, L1 retrotransposition can destabilize genomic DNA through a variety of different mechanisms (*e.g.*, insertional mutagenesis, shuffling genomic sequences, NHAR, *etc.*). Thus, it stands to reason that the cells may have adapted ways to control unabated L1 activity in order to preserve genomic DNA integrity. Some of the mechanisms that are thought to regulate L1 retrotransposition (Figure 1.6) will be detailed in this final section.

DNA methylation

The methylation of cytosine nucleotides in genomic DNA is a conserved gene silencing mechanism that can be found in many different organisms including plants, animals, and bacteria (Goll and Bestor 2005). In certain plant and animal species the majority of cytosine methylation is associated with transposable element sequences, which is thought to control their expression (Yoder et al. 1997, Martienssen 1998, Bestor 2003, Goll and Bestor 2005). Notably, DNA methylation is thought to function exclusively in the silencing of transposable elements in the fungus species, *Neurospora crassa* (Selker et al. 2003). Early L1 studies demonstrated a positive correlation between hypomethylation of the L1 promoter and the increased expression of L1 proteins in human cancer-derived cell lines (Bratthauer and Fanning 1993, Thayer et al. 1993, Alves et al. 1996), and that the methylation of CpG dinucleotides within the human L1 promoter represses L1 transcription (Hata and Sakaki 1997). More recent studies have shown that hypomethylation of L1 promoters is correlated with increase L1 expression in the human brain, and in select human induced pluripotent stem cells and hESC cell lines (Coufal et al. 2009, Wissing et al. 2012). Finally, L1 hypomethylation, increased L1 expression, and increased

retrotransposition activity have been detected in tumors from certain forms of human cancers (Suter et al. 2004, Iskow et al. 2010, Tubio et al. 2014).

During mammalian development, genomic DNA is successively de-methylated and re-methylated once in the early embryo and then once again in primordial germ cells. Once established in the early embryo, genomic DNA methylation patterns are thought to persist throughout development in adulthood. Notably, defects in establishing methylation of genomic DNA during development are linked to increased expression of L1 and other transposable elements and abnormal phenotypes in mutant mice. For example, loss of function mutations in DNA methyltransferase-1 (*Dnmt1*), a DNA methyltransferase responsible for maintaining DNA methylation during development, results in a dramatic increase in IAP LTR retrotransposon expression in mutant mouse embryos as well as embryonic lethality (Walsh et al. 1998). Similar phenotypes are observed in mutant mice with loss of function mutations in *Dnmt3L*, a gene that is necessary for establishing *de novo* methylation of genomic DNA in the germline. *Dnmt3L* mutant male mice exhibit a catastrophic defect in spermatogenesis that is accompanied by a loss of L1 and IAP element methylation and increased L1 and IAP element expression in germline tissue (Bourc'his and Bestor 2004). Notably, a recent study suggests that derepression of L1 sequences during epigenetic reprogramming of the embryonic germline may be linked to oocyte attrition in female mice (Malki et al. 2014). Despite these data, whether the derepression of L1 and other transposable element sequences plays a direct causative role in development defects will require further study.

Other mechanisms that silence L1 expression

In addition to DNA methylation, the modification of histone proteins (*e.g.*, acetylation, methylation) also may help regulate L1 expression. For example, one study using cultured human embryonic carcinoma cell lines suggests that histone deacetylase enzymes may act to maintain epigenetic silencing of newly integrated L1 insertions (Garcia-Perez et al. 2010). Histone modifications have also been suggested to regulate L1 expression in developing mouse embryos

and in germline cells (Di Giacomo et al. 2013, Fadloun et al. 2013, Pezic et al. 2014).

A number of cellular proteins that recognize epigenetic modifications in genomic DNA, such as methyl DNA binding proteins and zinc-finger proteins have been implicated in the regulation of L1 expression. For example, the methyl-CpG-binding protein 2 (MECP2) associates with L1 promoters and has been reported to repress L1 expression in neural stem cells and in transformed human cell lines (Yu et al. 2001, Muotri et al. 2010). Additionally, L1 activity is reported to be higher in the brains of *MeCP2* knockout mice and in the brains of patients with Rett Syndrome, a neurodevelopmental disorder typically caused by mutations in *MeCP2* (Muotri et al. 2010). Recent studies have also provided data suggesting that the several KRAB zinc-finger proteins as well as the promyelocytic Leukemia zinc finger protein (PLZF) may target genomic L1 and/or SVA elements for epigenetic silencing (Puszyk et al. 2013, Castro-Diaz et al. 2014, Jacobs et al. 2014).

Small RNA-mediated gene silencing

RNA interference (RNAi) is a highly conserved gene silencing mechanism in eukaryotes that is associated with gene regulation, imprinting, and silencing transposable elements. In RNAi, small silencing RNAs ranging between 21-31 nt are loaded onto Argonaute protein family members to form an RNA-induced silencing complex (RISC), which targets cellular mRNAs for translational inhibition and/or degradation. Two small RNA classes that target transposable elements include: small interfering RNAs (siRNAs) and PIWI-interacting RNAs (piRNAs).

siRNA: siRNAs are Dicer-dependent small RNAs (~21 nucleotide) that have been demonstrated to silence transposons in certain plant and invertebrate species (Saito and Siomi 2010, Castel and Martienssen 2013). Recent evidence suggests that siRNA may also silence transposable elements in mammals. For example, a recent study showed that most siRNAs in developing mouse oocytes correspond to retrotransposon sequences, and that LTR retrotransposon

transcripts are increased in oocytes from *Dicer* knockout mice (Watanabe et al. 2008). Finally, studies using cultured cells suggest that siRNAs may silence L1s in human tumor-derived cell lines (Soifer et al. 2005, Yang and Kazazian 2006) and possibly in mouse ESCs (Ciaudo et al. 2013).

piRNA and PIWI proteins: PIWI (p-element induced wimpy testis) - interacting RNAs (piRNAs) are short silencing RNAs (~26-31 nt) that function in a highly specialized RNA interference (RNAi) pathway that silences transposable elements in germline cells of several different animal species (Aravin et al. 2007, Siomi et al. 2011). In the piRNA pathway, transposable element piRNAs associate with PIWI proteins, which are germline specific argonaute protein family members to form the piRNA induced silencing complex (piRISC). The piRISC targets expressed transposable element mRNA sequences for destruction (Aravin et al. 2006, Girard et al. 2006, Vagin et al. 2006). A distinguishing feature of the piRNA pathway is the existence of a piRNA amplification loop called the ping-pong cycle, which generates additional piRNAs that target transposable elements (Brennecke et al. 2007, Gunawardane et al. 2007).

During the ping-pong cycle, once the piRISC cleaves transposable elements RNAs, the cleaved transposon RNAs, termed secondary piRNAs, are loaded onto a different PIWI protein family member to form a secondary piRISC (Siomi et al. 2011). The secondary piRISC is then thought to facilitate the processing of additional transposable element piRNAs, thus adding to the existing pool of piRISC complexes that target expressed transposable element RNAs. In principle, transposable element sequences that are most highly expressed (*i.e.*, most active) will preferentially generate additional piRISC complexes that target themselves. Thus, the piRNA ping-pong pathway is thought to serve as an adaptive defense mechanism against active transposable elements in the germline (Aravin et al. 2007, Siomi et al. 2011).

Several studies suggest that the piRNA pathway is required to silence transposable elements in the mammalian germline. In mice, a substantial

proportion of germline pre-pachytene piRNAs are derived from actively transcribed L1, SINE and LTR transposon sequences (Aravin et al. 2007, Aravin et al. 2008). Mouse studies have demonstrated a significant increase in L1 and IAP expression in the testis of mutant male mice deficient in the mouse PIWI protein, MILI (Piwi-Like RNA-Mediated Gene Silencing 2 (PIWIL2)) (Aravin et al. 2007). Notably, mutant male mice that lack functional mouse PIWI proteins, MILI and/or MIWI2 (Piwi-Like RNA-Mediated Gene Silencing 4 (PIWIL4)) display phenotypes reminiscent of those observed in male mice lacking functional *Dnmt3L* (Bourc'his and Bestor 2004) including sterility, massive de-repression of L1 and IAP elements and a corresponding demethylation of L1 sequences (Carmell et al. 2007, Aravin et al. 2008, Kuramochi-Miyagawa et al. 2008).

The loss in methylation of L1 and IAP element sequences in *MILI* and *MIWI2* mutant mice suggests that piRNAs also may be involved in establishing *de novo* methylation of genomic transposable elements sequences (Carmell et al. 2007, Aravin et al. 2008). Recent evidence suggests that mammalian piRNAs may also be involved in establishing repressive histone methylation to silence L1 expression in the germline (Di Giacomo et al. 2013, Pezic et al. 2014). Thus piRNAs seem to be involved in a coordinated defense against transposable elements that includes posttranscriptional destruction of transposable elements RNAs as well as epigenetic DNA modifications that silence transposable elements expression.

Posttranscriptional processing of L1 transcripts

Certain evidence suggests that L1 expression could be regulated by RNA splicing. Recent studies have demonstrated that L1 sequences encode a number of splice donor and acceptor sites that can result in extensive splicing of L1 RNA in human and mouse cells (Belancio et al. 2006, Belancio et al. 2008). Examination of various human tissues and human cultured cell lines has revealed that L1 RNA is expressed in a range of somatic tissues and that the level of L1 mRNA processing (*i.e.*, splicing) varies significantly among different human tissues and cell types (Belancio et al. 2010). These data suggest that

although L1 RNA may be normally expressed in certain adult somatic tissues, that posttranscriptional RNA splicing might limit the expression of full-length L1 RNA and thereby limit the level of L1 retrotransposition in somatic tissues.

RNA-binding proteins

Cytoplasmic RNA binding proteins also may function in the posttranscriptional regulation of L1 retrotransposition. A number of studies have demonstrated that L1 RNPs associate with a variety of cellular RNA binding proteins (Goodier et al. 2007, Dai et al. 2012, Goodier et al. 2013, Taylor et al. 2013). For example, immunofluorescence microscopy experiments have demonstrated that L1 ORF1p co-localizes with cytoplasmic stress granule (SG) proteins, including eIF3 (eukaryotic translation initiation factor 3), G3BP (GTPase Activating Protein (SH3 Domain) Binding Protein), and EIF2C2 (Argonaute 2 or Ago2) (Doucet et al. 2010, Goodier et al. 2010, Goodier et al. 2013). Stress granules are cytoplasmic RNP complexes that form in response to cellular stress and are thought to function in cytoplasmic mRNA metabolism and/or translation repression (Anderson and Kedersha 2008, Buchan and Parker 2009). Notably, SG and/or cytoplasmic processing body (PBs) formation is suppressed by some viruses, which suggests that SGs and PBs may serve a protective function during viral infections (Reineke and Lloyd 2013). Thus, it has been hypothesized that SGs and/or cellular RNA binding proteins may function to modulate L1 retrotransposition (Goodier et al. 2007).

Host antiviral proteins

Antiviral proteins typically function to restrict viral infectivity by recognizing specific viral components (*e.g.*, viral RNA and/or proteins) and blocking some aspect of the viral infectivity cycle (Goff 2004, Yan and Chen 2012, Zheng et al. 2012). A growing number of studies have demonstrated that host cell antiviral proteins also are capable of restricting L1 retrotransposition and the activity of other endogenous retrotransposons, suggesting a possible interrelationship

between the host cell response to viruses and the host cell response to transposable elements.

APOBEC3 proteins: In humans, there are seven *APOBEC3* (apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like3) genes that encode a small family of cytidine deaminases with RNA editing and/or DNA mutator activity (Chiu and Greene 2008). The finding that *APOBEC3G* (A3G) could restrict the activity of *vif*-deficient HIV-1 provided the first evidence that APOBEC3 proteins function as antiviral factors (Sheehy et al. 2002). Subsequent studies revealed that A3G also inhibits the activity of endogenous retroviruses and Alu elements suggesting that *APOBEC3* genes might also function to restrict the activity of endogenous retrotransposons. Subsequent characterization of *APOBEC3* genes by several labs has revealed that *APOBEC3A* (A3A), *APOBEC3B* (A3B), *APOBEC3C* (A3C) and *APOBEC3F* (A3F) inhibit L1 retrotransposition (Bogerd et al. 2006, Chen et al. 2006, Muckenfuss et al. 2006, Stenglein and Harris 2006, Richardson et al. 2014) and that A3A, A3B and A3G are able to inhibit Alu retrotransposition (Bogerd et al. 2006, Chiu et al. 2006).

APOBEC3 proteins may restrict retroviral activity through a deaminase-dependent mechanism, which involves editing of viral nucleic acid sequences, or by a deaminase-independent mechanism, which likely involves binding to viral RNA and blocking the procession of retroviral RT (Chiu and Greene 2008). Cell culture experiments have demonstrated that A3B and A3C mutants that lack enzymatic activity are able to inhibit L1 retrotransposition suggesting that some APOBEC3 proteins may inhibit L1 by a deaminase-independent mechanism (Bogerd et al. 2006, Horn et al. 2013). Notably, A3G-mediated restriction of Alu retrotransposition may also involve a deaminase-independent mechanism (Chiu et al. 2006). With this being stated, the localization of APOBEC3 proteins to the nucleus and the extensive editing of viral sequences suggest a cytidine deaminase-dependent mechanism could also apply to L1. (Bogerd et al. 2006). Notably, a recent study suggests that A3A may inhibit L1 retrotransposition by deaminating transiently exposed single-strand DNA during L1 integration

(Richardson et al. 2014) suggesting that deaminase activity may be required for A3A-mediated restriction.

Aicardi-Goutières syndrome (AGS) Genes: The three prime repair exonuclease 1 (TREX1) is the most abundant 3'-5' DNA exonuclease in cells. Mutations in the *TREX1* gene are associated with several autoimmune diseases including Aicardi-Goutières syndrome (AGS), familial chilblain lupus, and Systemic Lupus Erythematosus (SLE) (Kavanagh et al. 2008, Volkman and Stetson 2014). In a recent study, elevated levels of single stranded DNA from L1 and IAP elements was detected in TREX1 knockout mice suggesting that TREX1 might be involved in regulating endogenous retrotransposon activity. Experiments using cultured cells further demonstrated that overexpression of TREX1 restricts L1 retrotransposition and that TREX1 mutants deficient in catalytic activity were unable to mediate L1 inhibition (Stetson et al. 2008). Despite these observations how TREX1 restricts L1 retrotransposition remains unknown, but TREX1 restriction is thought to involve exonuclease activity.

Another cellular AGS gene that has been demonstrated to restrict the activity of L1 retrotransposition is SAM Domain and HD Domain 1 (*SAMHD1*) gene. Mutations in *SAMHD1* are associated with several human autoimmune disorders including AGS and SLE (Crow and Rehwinkel 2009). SAMHD1 was initially identified as an HIV-1 restriction factor in myeloid immune cells in screens to identify host cell factors that associate with the HIV-1 lentiviral auxiliary protein, *Vpx* (Hrecka et al. 2011, Laguette et al. 2011). The SAMHD1 protein catalyzes the hydrolysis of deoxynucleoside triphosphates (dNTP). SAMHD1 is thought to restrict retroviral activity in non-cycling cells by depleting cellular dNTP stores, which limits the supply of dNTPs available for the retroviral reverse transcription reaction (Ayinde et al. 2012). A recent study demonstrated that SAMHD1 restricts human and murine L1s in cell culture based retrotransposition assays. Interestingly, SAMHD1 mutants deficient in hydrolase activity were still able to inhibit L1 retrotransposition, suggesting that SAMHD1 may restrict retroviruses and L1 by different mechanisms (Zhao et al. 2013).

MOV10 and RNase L: Moloney leukemia virus 10 (MOV10) is an RNA helicase that has been demonstrated to inhibit the activity of HIV-1 and other retroviruses (Zheng et al. 2012, Gregersen et al. 2014). Several recent studies have reported that MOV10 can also restrict the activity of L1, Alu and mouse LTR retrotransposons (Arjan-Odedra et al. 2012, Goodier et al. 2012, Li et al. 2013). How MOV10 restricts retrotransposition is not known; however, evidence suggests that MOV10 binds to L1 RNA and may interfere with the expression of L1 ORF1p and/or with the reverse transcription of L1 cDNA (Goodier et al. 2012, Gregersen et al. 2014). Notably, recent evidence suggests that the MOV10 paralog, MOV10L, may also restrict the activity of endogenous retroelements in mice (Frost et al. 2010, Zheng et al. 2010). Finally, a recent study has demonstrated that the 2',5'-oligoadenylate (2-5A) synthetase (OAS)-RNase L system may also restrict endogenous retroelements (Zhang et al. 2014). In this study, the authors demonstrated that RNase L inhibits the activity of L1 and IAP elements in cultured cell assays. Mechanistic analyses showed that RNase L decreases the expression of L1 RNA and proteins suggesting that RNase L might directly target L1 RNA for degradation.

Closing Remarks

Barbara McClintock's discovery of transposons in maize occurred at a time before the discovery of the DNA double helix when genes were thought to be arranged neatly on chromosomes like beads-on-a-string (McClintock 1950, McClintock 1984). The discovery and continued study of transposable elements has revealed that genomes are not static, but are constantly changing and evolving. Notably, the finding that maize transposable elements could modify the expression of other maize genes was one of the first studied examples of a gene control system (Mcclintock 1961). For many years that followed these landmark discoveries, transposable elements remained underappreciated and perhaps were mostly thought of as rather useless junk. This outdated view has been challenged by recent scientific advancements such as the sequencing of the

human genome which have shown that active transposition can have a major impact on genomic DNA structure and function (Lander et al. 2001). Indeed, the initial discovery that L1 was active humans, the subsequent development of assays to test retrotransposition activity, and advancements in DNA sequencing technologies have enabled researchers to demonstrate that L1 retrotransposition continues to influence the human genome through mutational processes that can sometimes cause genetic diseases and/or generate genomic sequence changes that contribute to inter- and intra-individual diversity in the human population. Recent studies now suggest that somatic L1 retrotransposition may even influence the developing mammalian brain (Muotri et al. 2005, Coufal et al. 2009, Baillie et al. 2011, Evrony et al. 2012, Evrony et al. 2015, Upton et al. 2015). Notwithstanding, there are many important questions that should be addressed in future L1 studies such as how L1 integration is completed, how the host cell regulates L1 retrotransposition, and how if at all does L1 retrotransposition influence development of the mammalian brain. The main focus of this thesis is to further explore the question of what host cell factors are involved in regulating L1 retrotransposition.

Image courtesy of Dr. Gerry Neuffer

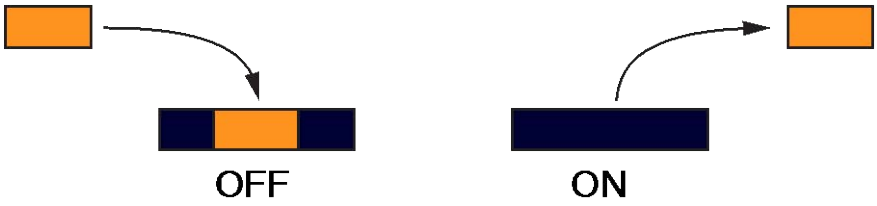


Figure 1.1: Transposable elements and corn kernel variegation in maize.

Top panel: Image of variegated (*i.e.*, purple dots on colorless white background) corn kernels. The image is reproduced with permission from Dr. Gerry Neuffer and the maize genetics and genomics database (Lawrence et al. 2004). *Bottom panel:* The insertion of a transposable element (orange rectangle) into a gene that controls corn kernel color (dark purple rectangle) turns off gene expression resulting in colorless corn kernels (colorless corn kernels in top panel). If the transposable element transposes out of the gene during corn kernel development, gene expression is turned back on giving rise to colored sectors or spots on a colorless background. Large spots result from transposition occurring early in development and small spots result from transposition occurring late in development.

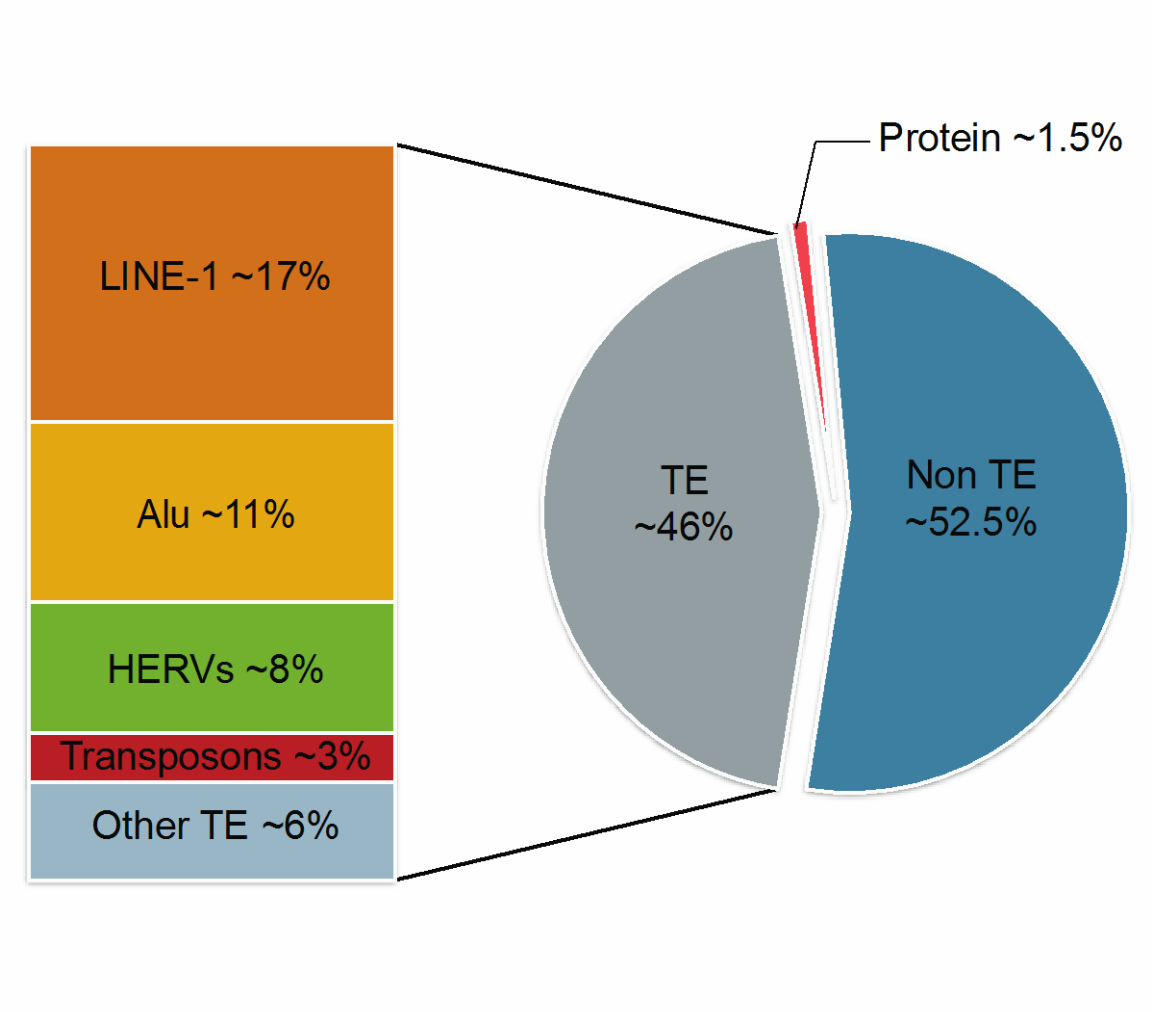


Figure 1.2: Transposable element content of the human genome.

Approximately 46% of the human genome is comprised of transposable element (TE) sequences. Only ~1.5% of the genome is known to encode proteins. The distribution (% of total human genome) of the major types of human TE is indicated to the left of the pie chart (Lander et al. 2001).

Figure 1.3: Schematic diagram of the major transposable elements in the human genome.

A) Diagram of a human DNA transposon. Human DNA transposons are ~2-3 kb in length and typically encode a transposase enzyme (orange rectangle) that is flanked by short inverted tandem repeat sequences (white triangles). B) Diagram of a human LTR retrotransposon (human endogenous retrovirus or HERV). HERVs are flanked by long tandem repeats (LTRs) that surround structural and reverse transcriptase genes similar to retroviral *gag* and *pol* genes respectively. Some HERVs may also encode a non-functional envelope (*env*) gene. (C-E) Human non-LTR retrotransposons C) Schematic of an autonomous human L1 element. L1 is ~6 kb in length and encodes a 5' UTR that promotes L1 transcription. The 5' untranslated region (5' UTR) is followed by two ORFs that are separated by a 63 bp intergenic spacer and a short 3' UTR that terminates in a variable length poly adenosine tract. ORF1 encodes an ~40 kDa nucleic acid binding protein (ORF1p) and ORF2 encodes ORF2p, an ~150 kDa protein with RT and EN activities. Both L1 proteins are required for L1 retrotransposition. D) Schematic diagram of a human Alu element. Alu is non-autonomous and therefore requires the L1-encoded proteins in order to retrotranspose. Human Alu elements are ~300 bp in length and exhibit a dimeric sequence arrangement that consists of diverged left and right monomers (dark green and light green rectangles respectively) that are derived from 7SL RNA and are separated by a short poly adenosine tract. Alu elements end in a variable length poly adenosine tail. E) Diagram of a processed pseudogene. Processed pseudogenes are non-functional genes that result from the retrotransposition of cellular mRNAs by the L1-encoded protein machinery. Processed pseudogenes resemble cellular mRNAs and typically are comprised of 5' and 3' UTRs (grey rectangles), coding exons (blue boxes with numbers indicating individual exons (1-3 in this example)), and a variable length poly adenosine tail (see main text for references).

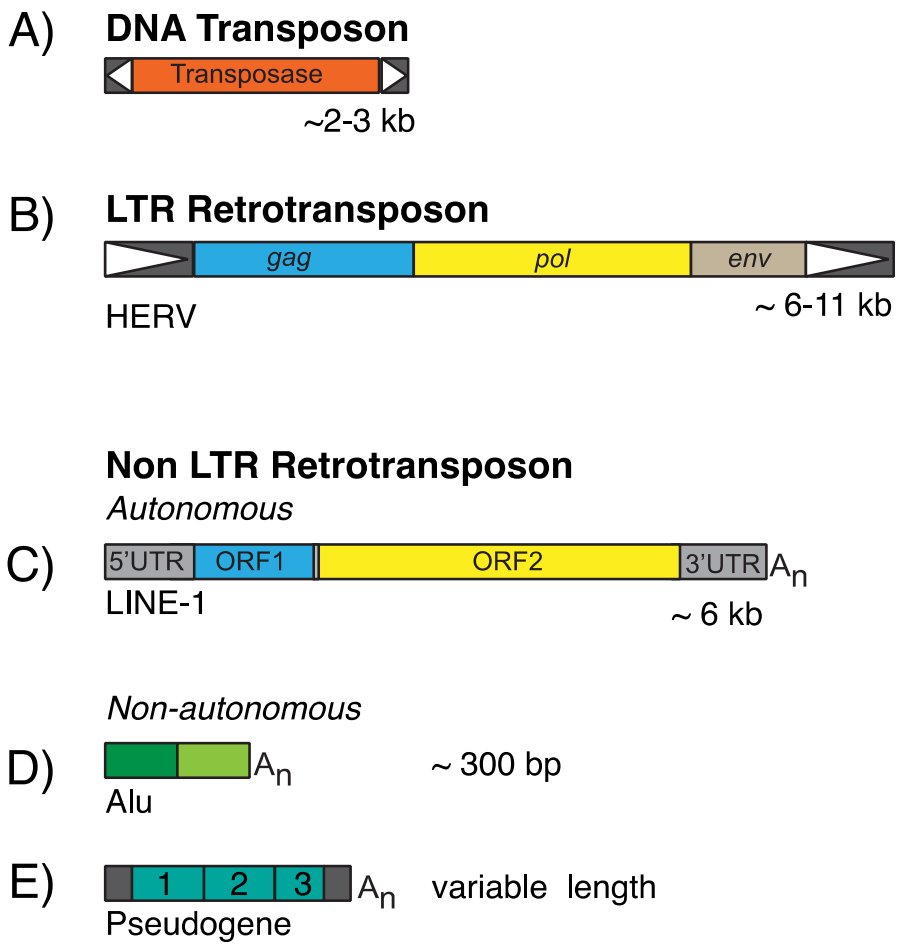


Figure 1.3: Schematic diagram of the major transposable elements in the human genome.

Figure 1.4: Diagram of the L1 retrotransposition cycle.

Following transcription of a genomic L1 element (red rectangle), the resulting L1 mRNA is exported to the cytoplasm. The L1 mRNA is translated in the cytoplasm to produce ORF1p (blue circle) and ORF2p (yellow circle), which preferentially bind to the L1 mRNA (*cis* preference) to form a L1 ribonucleoprotein particle (L1 RNP). The L1-encoded proteins may also act in *trans* to retrotranspose Alu (green RNA structure) or other cellular RNAs (*e.g.*, SVA, mRNA, *etc.*). Once components of the L1 RNP enter the nucleus, a new L1 copy is integrated into genomic DNA by the process of target site primed reverse transcription (TPRT). For L1 TPRT, ORF2p EN makes a single-strand nick in genomic DNA at a degenerate consensus sequence (5'-TTTT/A-3', where "/" denotes the nick) to expose a free 3'-OH that can be used by the ORF2p RT as a primer to initiate L1 reverse transcription. The L1 ORF2p RT uses the L1 mRNA as a template to synthesize an antisense L1 cDNA copy (red line) directly onto the exposed genomic DNA strand. How the remaining steps of L1 integration occur requires further investigation, but may involve other host cell proteins such as DNA repair enzymes.

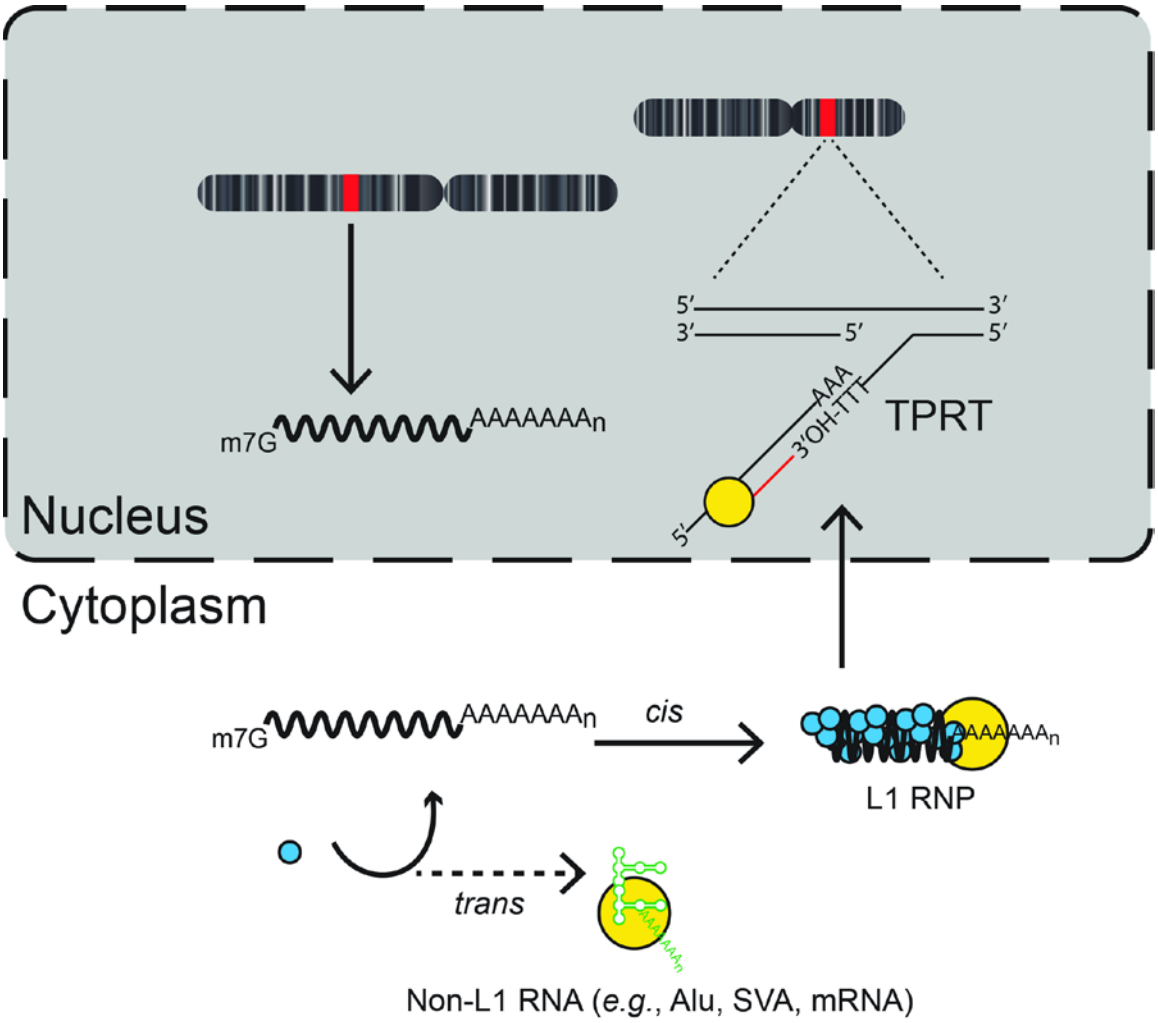


Figure 1.4: Diagram of the L1 retrotransposition cycle.

Figure 1.5: Examples of how L1 impacts human genomic DNA.

A) L1 (blue rectangle) can insert into a gene (orange rectangle) and disrupt normal gene expression. B) L1 insertions can result in deletions of genomic DNA (black rectangle represents genomic DNA sequence that is deleted). C) L1 can retrotranspose genomic DNA sequences that lie adjacent to the 5' end (a 5' transduction; light grey rectangle) or the 3' end (3' transduction; black rectangle) of the L1 element. D) L1 encoded proteins (blue circle) can mobilize Alu (green rectangle) and/or other cellular RNAs (*e.g.*, SVA, mRNA, *etc.*) in *trans*. E) A L1 insertion into gene coding region such as an intron (grey rectangles; exons are yellow, orange, green rectangles numbered 1-3 respectively) could result in premature polyadenylation and produce a truncated RNA transcript (yellow/orange rectangle), or a missplicing event resulting in an RNA transcript that is missing an exon (yellow/green rectangle) F) Post-insertional illegitimate recombination event (zigzag arrowed line) between non-allelic L1 sequences (blue rectangles) could result in the duplication (bottom left) or deletion (bottom right) of genomic DNA segments (grey rectangle).

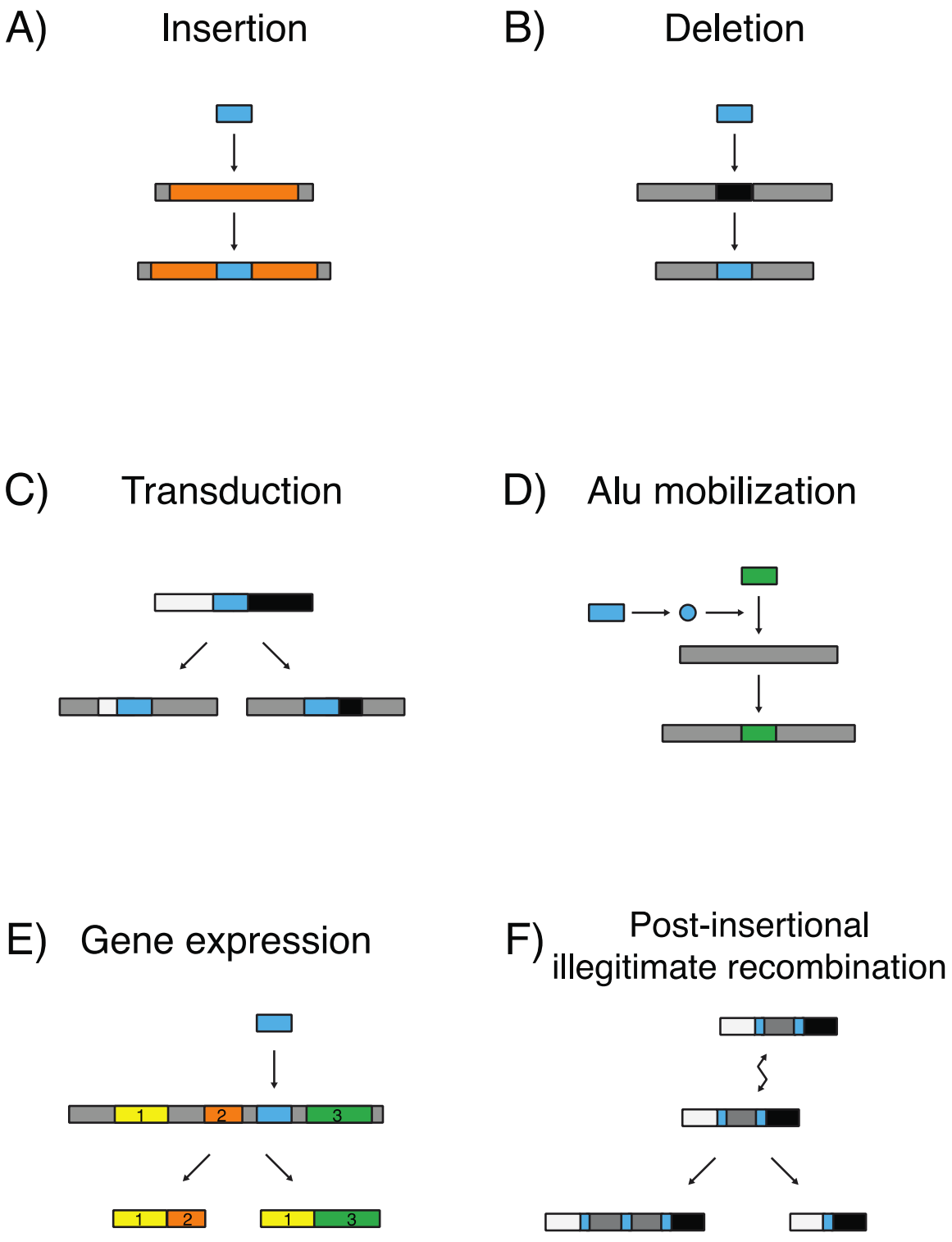


Figure 1.5: Examples of how L1 impacts human genomic DNA.

Figure 1.6: Host cell mechanisms that restrict L1 activity.

Host cell mechanisms may function to regulate several phases (numbered 1-5) of the L1 retrotransposition cycle. 1) Epigenetic modification of genomic L1 sequences such as cytosine methylation at L1 promoters is thought to limit L1 expression. 2) Expressed L1 RNA may be subject to alternative splicing and or regulation by splicing factors such as hnRNPL, which limits expression of full-length L1 RNA. The piRNA pathway may target the destruction of L1 transcripts in germline cells. 3) Host cell antiviral proteins (*e.g.*, MOV10, RNase L) may prevent the accumulation of L1 RNA and/or proteins in the cytoplasm. 4) Cytoplasmic stress granules (SGs) may function to sequester components of the L1 RNP (*i.e.*, L1 RNA and/or L1-encoded proteins) in the cytoplasm. 5) Certain host antiviral proteins (*e.g.*, APOBEC3A (A3A), SAMHD1) also may target the production of L1 cDNA by various mechanisms (see main text).

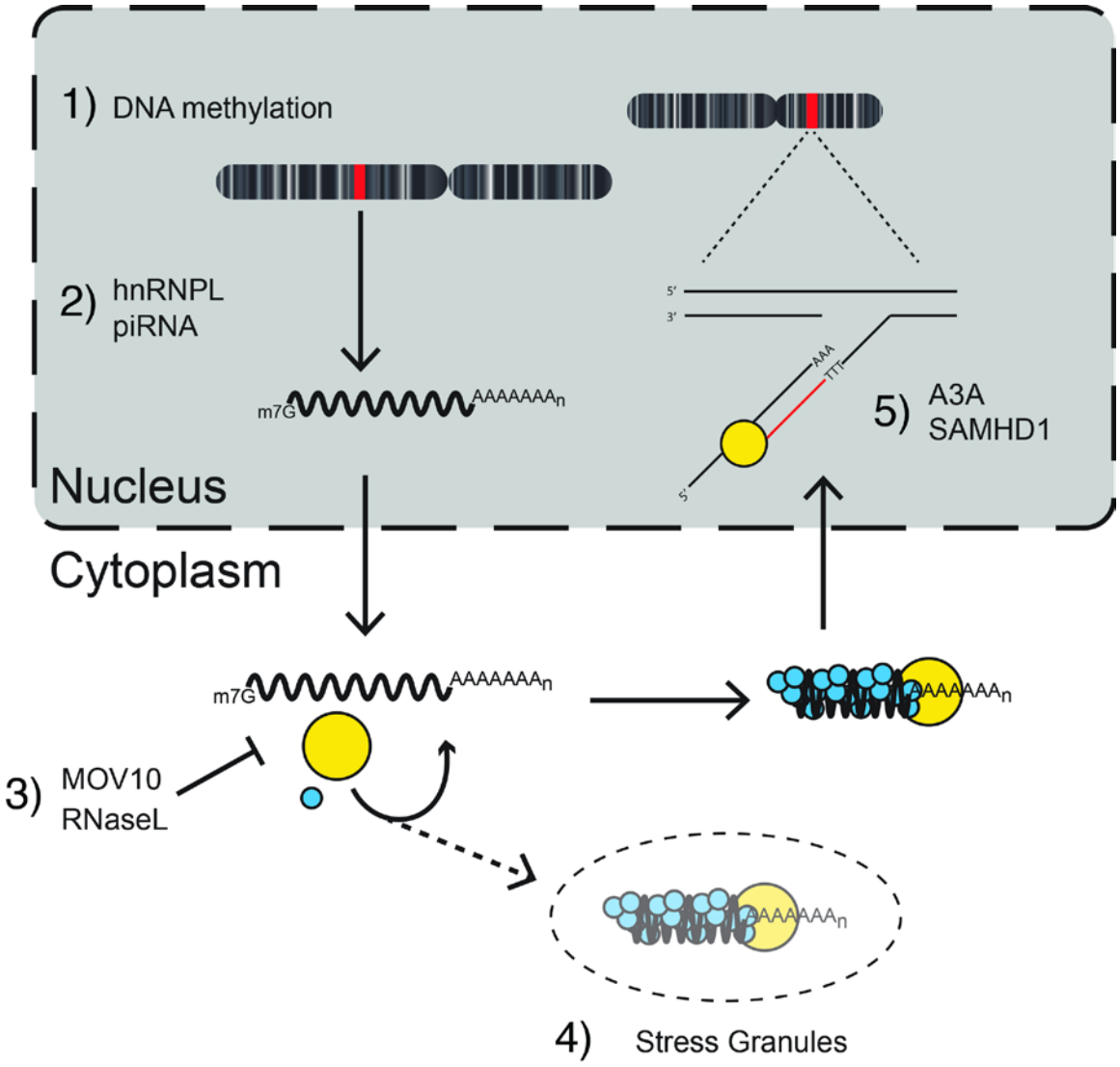


Figure 1.6: Host cell mechanisms that restrict L1 activity.

References

Alisch, R. S., J. L. Garcia-Perez, A. R. Muotri, F. H. Gage and J. V. Moran, "Unconventional translation of mammalian LINE-1 retrotransposons." Genes Dev, 2006 **20**(2): 210-224.

Alves, G., A. Tatro and T. Fanning, "Differential methylation of human LINE-1 retrotransposons in malignant cells." Gene, 1996 **176**(1-2): 39-44.

An, W., J. S. Han, S. J. Wheelan, E. S. Davis, C. E. Coombes, P. Ye, C. Triplett and J. D. Boeke, "Active retrotransposition by a synthetic L1 element in mice." Proc Natl Acad Sci U S A, 2006 **103**(49): 18662-18667.

Anderson, P. and N. Kedersha, "Stress granules: the Tao of RNA triage." Trends Biochem Sci, 2008 **33**(3): 141-150.

Aravin, A., D. Gaidatzis, S. Pfeffer, M. Lagos-Quintana, P. Landgraf, N. Iovino, P. Morris, M. J. Brownstein, S. Kuramochi-Miyagawa, T. Nakano, M. Chien, J. J. Russo, J. Ju, R. Sheridan, C. Sander, M. Zavolan and T. Tuschl, "A novel class of small RNAs bind to MILI protein in mouse testes." Nature, 2006 **442**(7099): 203-207.

Aravin, A. A., G. J. Hannon and J. Brennecke, "The Piwi-piRNA Pathway Provides an Adaptive Defense in the Transposon Arms Race." Science, 2007 **318**(5851): 761-764.

Aravin, A. A., R. Sachidanandam, D. Bourc'his, C. Schaefer, D. Pezic, K. F. Toth, T. Bestor and G. J. Hannon, "A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice." Mol Cell, 2008 **31**(6): 785-799.

Aravin, A. A., R. Sachidanandam, A. Girard, K. Fejes-Toth and G. J. Hannon, "Developmentally regulated piRNA clusters implicate MILI in transposon control." Science, 2007 **316**(5825): 744-747.

Arjan-Odedra, S., C. M. Swanson, N. M. Sherer, S. M. Wolinsky and M. H. Malim, "Endogenous MOV10 inhibits the retrotransposition of endogenous retroelements but not the replication of exogenous retroviruses." Retrovirology, 2012 **9**: 53.

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): 3846-3855.

Ayinde, D., N. Casartelli and O. Schwartz, "Restricting HIV the SAMHD1 way: through nucleotide starvation." Nat Rev Microbiol, 2012 **10**(10): 675-680.

Bailey, J. A., Z. Gu, R. A. Clark, K. Reinert, R. V. Samonte, S. Schwartz, M. D. Adams, E. W. Myers, P. W. Li and E. E. Eichler, "Recent segmental duplications in the human genome." Science, 2002 **297**(5583): 1003-1007.

Bailey, J. A., G. Liu and E. E. Eichler, "An Alu transposition model for the origin and expansion of human segmental duplications." Am J Hum Genet, 2003 **73**(4): 823-834.

Baillie, J. K., M. W. Barnett, K. R. Upton, D. J. Gerhardt, T. A. Richmond, F. De Sapio, P. M. Brennan, P. Rizzu, S. Smith, M. Fell, R. T. Talbot, S. Gustincich, T. C. Freeman, J. S. Mattick, D. A. Hume, P. Heutink, P. Carninci, J. A. Jeddeloh and G. J. Faulkner, "Somatic retrotransposition alters the genetic landscape of the human brain." Nature, 2011 **479**(7374): 534-537.

Bannert, N. and R. Kurth, "The evolutionary dynamics of human endogenous retroviral families." Annual review of genomics and human genetics, 2006 **7**: 149-173.

Basame, S., P. Wai-Lun Li, G. Howard, D. Branciforte, D. Keller and S. L. Martin, "Spatial Assembly and RNA Binding Stoichiometry of a LINE-1 Protein Essential for Retrotransposition." J Mol Biol, 2006.

Beauregard, A., M. J. Curcio and M. Belfort, "The take and give between retrotransposable elements and their hosts." Annu Rev Genet, 2008 **42**: 587-617.

Beck, C. R., P. Collier, C. Macfarlane, M. Malig, J. M. Kidd, E. E. Eichler, R. M. Badge and J. V. Moran, "LINE-1 retrotransposition activity in human genomes." Cell, 2010 **141**(7): 1159-1170.

Beck, C. R., J. L. Garcia-Perez, R. M. Badge and J. V. Moran, "LINE-1 elements in structural variation and disease." Annu Rev Genomics Hum Genet, 2011 **12**: 187-215.

Becker, K. G., G. D. Swergold, K. Ozato and R. E. Thayer, "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): 1697-1702.

Belancio, V. P., D. J. Hedges and P. Deininger, "LINE-1 RNA splicing and influences on mammalian gene expression." Nucleic Acids Research, 2006 **34**(5): 1512-1521.

Belancio, V. P., A. M. Roy-Engel and P. Deininger, "The impact of multiple splice sites in human L1 elements." Gene, 2008 **411**(1-2): 38-45.

Belancio, V. P., A. M. Roy-Engel, R. R. Pochampally and P. Deininger, "Somatic expression of LINE-1 elements in human tissues." Nucleic Acids Res, 2010 **38**(12): 3909-3922.

Bestor, T. H., "Cytosine methylation mediates sexual conflict." Trends Genet, 2003 **19**(4): 185-190.

Boeke, J. D., D. J. Garfinkel, C. A. Styles and G. R. Fink, "Ty elements transpose through an RNA intermediate." Cell, 1985 **40**(3): 491-500.

Bogerd, H. P., H. L. Wiegand, A. E. Hulme, J. L. Garcia-Perez, K. S. O'Shea, J. V. Moran and B. R. Cullen, "Cellular inhibitors of long interspersed element 1 and Alu retrotransposition." Proc Natl Acad Sci U S A, 2006 **103**(23): 8780-8785.

Bourc'his, D. and T. H. Bestor, "Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L." Nature, 2004 **431**(7004): 96-99.

Bradley, J., A. Baltus, H. Skaletsky, M. Royce-Tolland, K. Dewar and D. C. Page, "An X-to-autosome retrogene is required for spermatogenesis in mice." Nat Genet, 2004 **36**(8): 872-876.

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): 2584-2592.

Bratthauer, G. L. and T. G. Fanning, "LINE-1 retrotransposon expression in pediatric germ cell tumors." Cancer, 1993 **71**(7): 2383-2386.

Brennan, G., Y. Kozyrev and S.-L. Hu, "TRIMCyp expression in Old World primates Macaca nemestrina and Macaca fascicularis." Proceedings of the National Academy of Sciences, 2008 **105**(9): 3569-3574.

Brennecke, J., A. A. Aravin, A. Stark, M. Dus, M. Kellis, R. Sachidanandam and G. J. Hannon, "Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila." Cell, 2007 **128**(6): 1089-1103.

Brooks, M., W. Gu, J. Barnas, J. Ray and K. Ray, "A Line 1 insertion in the Factor IX gene segregates with mild hemophilia B in dogs." Mammalian Genome, 2003 **14**(11): 788-795.

Brouha, B., C. Meischl, E. Ostertag, M. de Boer, Y. Zhang, H. Neijens, D. Roos and H. H. Kazazian, Jr., "Evidence consistent with human L1 retrotransposition in maternal meiosis I." Am J Hum Genet, 2002 **71**(2): 327-336.

Brouha, B., J. Schustak, R. M. Badge, S. Lutz-Prigge, A. H. Farley, J. V. Moran and H. H. Kazazian, Jr., "Hot L1s account for the bulk of retrotransposition in the human population." Proc Natl Acad Sci U S A, 2003 **100**(9): 5280-5285.

Buchan, J. R. and R. Parker, "Eukaryotic Stress Granules: The Ins and Outs of Translation." Molecular Cell, 2009 **36**(6): 932-941.

Burgess, D. L., D. C. Kohrman, J. Galt, N. W. Plummer, J. M. Jones, B. Spear and M. H. Meisler, "Mutation of a new sodium channel gene, Scn8a, in the mouse mutant 'motor endplate disease'." Nature genetics, 1995 **10**(4): 461-465.

Burwinkel, B. and M. W. Kilimann, "Unequal homologous recombination between LINE-1 elements as a mutational mechanism in human genetic disease." J Mol Biol, 1998 **277**(3): 513-517.

Buzdin, A., S. Ustyugova, E. Gogvadze, T. Vinogradova, Y. Lebedev and E. Sverdlov, "A new family of chimeric retrotranscripts formed by a full copy of U6 small nuclear RNA fused to the 3' terminus of l1." Genomics, 2002 **80**(4): 402-406.

Callahan, K. E., A. B. Hickman, C. E. Jones, R. Ghirlando and A. V. Furano, "Polymerization and nucleic acid-binding properties of human L1 ORF1 protein." Nucleic Acids Res, 2012 **40**(2): 813-827.

Carmell, M. A., A. l. Girard, H. J. G. van de Kant, D. Bourc'his, T. H. Bestor, D. G. de Rooij and G. J. Hannon, "MIWI2 Is Essential for Spermatogenesis and

Repression of Transposons in the Mouse Male Germline." Developmental Cell, 2007 **12**(4): 503-514.

Castel, S. E. and R. A. Martienssen, "RNA interference in the nucleus: roles for small RNAs in transcription, epigenetics and beyond." Nat Rev Genet, 2013 **14**(2): 100-112.

Castro-Diaz, N., G. Ecco, A. Coluccio, A. Kapopoulou, B. Yazdanpanah, M. Friedli, J. Duc, S. M. Jang, P. Turelli and D. Trono, "Evolutionally dynamic L1 regulation in embryonic stem cells." Genes & Development, 2014 **28**(13): 1397-1409.

Chen, H., C. E. Lilley, Q. Yu, D. V. Lee, J. Chou, I. Narvaiza, N. R. Landau and M. D. Weitzman, "APOBEC3A is a potent inhibitor of adeno-associated virus and retrotransposons." Curr Biol, 2006 **16**(5): 480-485.

Chen, J., A. Rattner and J. Nathans, "Effects of L1 retrotransposon insertion on transcript processing, localization and accumulation: lessons from the retinal degeneration 7 mouse and implications for the genomic ecology of L1 elements." Hum Mol Genet, 2006 **15**(13): 2146-2156.

Chen, J. M., P. D. Stenson, D. N. Cooper and C. Ferec, "A systematic analysis of LINE-1 endonuclease-dependent retrotranspositional events causing human genetic disease." Hum Genet, 2005 **117**(5): 411-427.

Chiu, Y.-L., H. E. Witkowska, S. C. Hall, M. Santiago, V. B. Soros, C. c. Esnault, T. Heidmann and W. C. Greene, "High-molecular-mass APOBEC3G complexes restrict Alu retrotransposition." Proceedings of the National Academy of Sciences, 2006 **103**(42): 15588-15593.

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**: 317-353.

Chu, W. M., W. M. Liu and C. W. Schmid, "RNA polymerase III promoter and terminator elements affect Alu RNA expression." Nucleic Acids Research, 1995 **23**(10): 1750-1757.

Ciaudo, C., F. Jay, I. Okamoto, C. J. Chen, A. Sarazin, N. Servant, E. Barillot, E. Heard and O. Voinnet, "RNAi-dependent and independent control of LINE1 accumulation and mobility in mouse embryonic stem cells." PLoS Genet, 2013 **9**(11): e1003791.

Clements, A. P. and M. F. Singer, "The human LINE-1 reverse transcriptase: Effect of deletions outside the common reverse transcriptase domain." Nucleic Acids Research, 1998 **26**(15): 3528-3535.

Conrad, B., R. N. Weissmahr, J. Boni, R. Arcari, J. Schupbach and B. Mach, "A human endogenous retroviral superantigen as candidate autoimmune gene in type I diabetes." Cell, 1997 **90**(2): 303-313.

Cordaux, R. and M. A. Batzer, "The impact of retrotransposons on human genome evolution." Nat Rev Genet, 2009 **10**(10): 691-703.

Cost, G. J., A. Golding, M. S. Schlissel and J. D. Boeke, "Target DNA chromatinization modulates nicking by L1 endonuclease." Nucleic Acids Res, 2001 **29**(2): 573-577.

Coufal, N. G., J. L. Garcia-Perez, G. E. Peng, G. W. Yeo, Y. Mu, M. T. Lovci, M. Morell, K. S. O'Shea, J. V. Moran and F. H. Gage, "L1 retrotransposition in human neural progenitor cells." Nature, 2009 **460**(7259): 1127-1131.

Craig, N. L., R. Craigie, M. Gellert and A. M. Lambowitz, "Mobile DNA II. American Society for Microbiology Press." Washington, DC, 2002.

Crow, Y. J. and J. Rehwinkel, "Aicardi-Goutieres syndrome and related phenotypes: linking nucleic acid metabolism with autoimmunity." Hum Mol Genet, 2009 **18**(R2): R130-136.

Dai, L., M. S. Taylor, K. A. O'Donnell and J. D. Boeke, "Poly(A) binding protein C1 is essential for efficient L1 retrotransposition and affects L1 RNP formation." Mol Cell Biol, 2012 **32**(21): 4323-4336.

Damert, A., J. Raiz, A. V. Horn, J. Lower, H. Wang, J. Xing, M. A. Batzer, R. Lower and G. G. Schumann, "5'-Transducing SVA retrotransposon groups spread efficiently throughout the human genome." Genome Res, 2009 **19**(11): 1992-2008.

de Koning, A. P. J., W. Gu, T. A. Castoe, M. A. Batzer and D. D. Pollock, "Repetitive elements may comprise over two-thirds of the human genome." PLoS genetics, 2011 **7**(12): e1002384.

Deininger, P. L., D. J. Jolly, C. M. Rubin, T. Friedmann and C. W. Schmid, "Base sequence studies of 300 nucleotide renatured repeated human DNA clones." Journal of Molecular Biology, 1981 **151**(1): 17-33.

Dewannieux, M., C. Esnault and T. Heidmann, "LINE-mediated retrotransposition of marked Alu sequences." Nat Genet, 2003 **35**(1): 41-48.

Di Giacomo, M., S. Comazzetto, H. Saini, S. De Fazio, C. Carrieri, M. Morgan, L. Vasiliauskaite, V. Benes, A. J. Enright and D. O'Carroll, "Multiple epigenetic mechanisms and the piRNA pathway enforce LINE1 silencing during adult spermatogenesis." Mol Cell, 2013 **50**(4): 601-608.

Dmitriev, S. E., D. E. Andreev, I. M. Terenin, I. A. Olovnikov, V. S. Prassolov, W. C. Merrick and I. N. Shatsky, "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): 4685-4697.

Dombroski, B. A., Q. Feng, S. L. Mathias, D. M. Sassaman, A. F. Scott, H. H. Kazazian, Jr. and J. D. Boeke, "An in vivo assay for the reverse transcriptase of human retrotransposon L1 in Saccharomyces cerevisiae." Mol Cell Biol, 1994 **14**(7): 4485-4492.

Dombroski, B. A., S. L. Mathias, E. Nanthakumar, A. F. Scott and H. H. Kazazian, Jr., "Isolation of an active human transposable element." Science, 1991 **254**(5039): 1805-1808.

Doolittle, W. F. and C. Sapienza, "Selfish genes, the phenotype paradigm and genome evolution." Nature, 1980 **284**(5757): 601-603.

Doucet, A. J., A. E. Hulme, E. Sahinovic, D. A. Kulpa, J. B. Moldovan, H. C. Kopera, J. N. Athanikar, M. Hasnaoui, A. Bucheton, J. V. Moran and N. Gilbert, "Characterization of LINE-1 ribonucleoprotein particles." PLoS Genet, 2010 **6**(10): e1001150.

Eickbush, T. H., "Telomerase and retrotransposons: which came first?" Science, 1997 **277**(5328): 911-912.

Ergun, S., C. Buschmann, J. Heukeshoven, K. Dammann, F. Schnieders, H. Lauke, F. Chalajour, N. Kilic, W. H. Stratling and G. G. Schumann, "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): 27753-27763.

Esnault, C., J. Maestre and T. Heidmann, "Human LINE retrotransposons generate processed pseudogenes." Nat Genet, 2000 **24**(4): 363-367.

Evrony, G. D., X. Cai, E. Lee, L. B. Hills, P. C. Elhosary, H. S. Lehmann, J. J. Parker, K. D. Atabay, E. C. Gilmore, A. Poduri, P. J. Park and C. A. Walsh, "Single-neuron sequencing analysis of L1 retrotransposition and somatic mutation in the human brain." Cell, 2012 **151**(3): 483-496.

Evrony, G. D., E. Lee, B. K. Mehta, Y. Benjamini, R. M. Johnson, X. Cai, L. Yang, P. Haseley, H. S. Lehmann, P. J. Park and C. A. Walsh, "Cell lineage analysis in human brain using endogenous retroelements." Neuron, 2015 **85**(1): 49-59.

Ewing, A. D. and H. H. Kazazian, Jr., "High-throughput sequencing reveals extensive variation in human-specific L1 content in individual human genomes." Genome Res, 2010 **20**(9): 1262-1270.

Fadloun, A., S. p. Le Gras, B. Jost, C. l. Ziegler-Birling, H. Takahashi, E. Gorab, P. Carninci and M.-E. Torres-Padilla, "Chromatin signatures and retrotransposon profiling in mouse embryos reveal regulation of LINE-1 by RNA." Nat Struct Mol Biol, 2013 **advance online publication**.

Fanning, T. and M. Singer, "The LINE-1 DNA sequences in four mammalian orders predict proteins that conserve homologies to retrovirus proteins." Nucleic Acids Res, 1987 **15**(5): 2251-2260.

Faulkner, G. J., Y. Kimura, C. O. Daub, S. Wani, C. Plessy, K. M. Irvine, K. Schroder, N. Cloonan, A. L. Steptoe, T. Lassmann, K. Waki, N. Hornig, T. Arakawa, H. Takahashi, J. Kawai, A. R. Forrest, H. Suzuki, Y. Hayashizaki, D. A. Hume, V. Orlando, S. M. Grimmond and P. Carninci, "The regulated retrotransposon transcriptome of mammalian cells." Nat Genet, 2009 **41**(5): 563-571.

Feng, Q., J. V. Moran, H. H. Kazazian, Jr. and J. D. Boeke, "Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition." Cell, 1996 **87**(5): 905-916.

Fitch, D. H., W. J. Bailey, D. A. Tagle, M. Goodman, L. Sieu and J. L. Slightom, "Duplication of the gamma-globin gene mediated by L1 long interspersed repetitive elements in an early ancestor of simian primates." Proc Natl Acad Sci U S A, 1991 **88**(16): 7396-7400.

Frost, R. J. A., F. K. Hamra, J. A. Richardson, X. Qi, R. Bassel-Duby and E. N. Olson, "MOV10L1 is necessary for protection of spermatocytes against retrotransposons by Piwi-interacting RNAs." Proceedings of the National Academy of Sciences, 2010 **107**(26): 11847-11852.

Galli, U. M., M. Sauter, B. Lecher, S. Maurer, H. Herbst, K. Roemer and N. Mueller-Lantzsch, "Human endogenous retrovirus rec interferes with germ cell development in mice and may cause carcinoma in situ, the predecessor lesion of germ cell tumors." Oncogene, 2005 **24**(19): 3223-3228.

Garcia-Perez, J. L., A. J. Doucet, A. Bucheton, J. V. Moran and N. Gilbert, "Distinct mechanisms for trans-mediated mobilization of cellular RNAs by the LINE-1 reverse transcriptase." Genome Res, 2007 **17**(5): 602-611.

Garcia-Perez, J. L., M. C. N. Marchetto, A. R. Muotri, N. G. Coufal, F. H. Gage, K. S. O'Shea and J. V. Moran, "LINE-1 retrotransposition in human embryonic stem cells." Human Molecular Genetics, 2007 **16**(13): 1569-1577.

Garcia-Perez, J. L., M. Morell, J. O. Scheys, D. A. Kulpa, S. Morell, C. C. Carter, G. D. Hammer, K. L. Collins, K. S. O/'Shea, P. Menendez and J. V. Moran, "Epigenetic silencing of engineered L1 retrotransposition events in human embryonic carcinoma cells." Nature, 2010 **466**(7307): 769-773.

Georgiou, I., D. Noutsopoulos, E. Dimitriadou, G. Markopoulos, A. Apergi, L. Lazaros, T. Vaxevanoglou, K. Pantos, M. Syrrou and T. Tzavaras, "Retrotransposon RNA expression and evidence for retrotransposition events in human oocytes." Hum Mol Genet, 2009 **18**(7): 1221-1228.

Gibbs, R. A., G. M. Weinstock, M. L. Metzker, D. M. Muzny, E. J. Sodergren, S. Scherer, G. Scott, D. Steffen, K. C. Worley, P. E. Burch, G. Okwuonu, S. Hines, L. Lewis, C. DeRamo, O. Delgado, S. Dugan-Rocha, G. Miner, M. Morgan, A. Hawes, R. Gill, Celera, R. A. Holt, M. D. Adams, P. G. Amanatides, H. Baden-Tillson, M. Barnstead, S. Chin, C. A. Evans, S. Ferriera, C. Fosler, A. Glodek, Z. Gu, D. Jennings, C. L. Kraft, T. Nguyen, C. M. Pfannkoch, C. Sitter, G. G. Sutton, J. C. Venter, T. Woodage, D. Smith, H. M. Lee, E. Gustafson, P. Cahill, A. Kana, L. Doucette-Stamm, K. Weinstock, K. Fechtel, R. B. Weiss, D. M. Dunn, E. D. Green, R. W. Blakesley, G. G. Bouffard, P. J. De Jong, K. Osoegawa, B. Zhu, M. Marra, J. Schein, I. Bosdet, C. Fjell, S. Jones, M. Krzywinski, C. Mathewson, A. Siddiqui, N. Wye, J. McPherson, S. Zhao, C. M. Fraser, J. Shetty, S. Shatsman, K. Geer, Y. Chen, S. Abramzon, W. C. Nierman, P. H. Havlak, R. Chen, K. J. Durbin, A. Egan, Y. Ren, X. Z. Song, B. Li, Y. Liu, X. Qin, S. Cawley, K. C. Worley, A. J. Cooney, L. M. D'Souza, K. Martin, J. Q. Wu, M. L. Gonzalez-Garay,

A. R. Jackson, K. J. Kalafus, M. P. McLeod, A. Milosavljevic, D. Virk, A. Volkov, D. A. Wheeler, Z. Zhang, J. A. Bailey, E. E. Eichler, E. Tuzun, E. Birney, E. Mongin, A. Ureta-Vidal, C. Woodwark, E. Zdobnov, P. Bork, M. Suyama, D. Torrents, M. Alexandersson, B. J. Trask, J. M. Young, H. Huang, H. Wang, H. Xing, S. Daniels, D. Gietzen, J. Schmidt, K. Stevens, U. Vitt, J. Wingrove, F. Camara, M. Mar Alba, J. F. Abril, R. Guigo, A. Smit, I. Dubchak, E. M. Rubin, O. Couronne, A. Poliakov, N. Hubner, D. Ganten, C. Goesele, O. Hummel, T. Kreitler, Y. A. Lee, J. Monti, H. Schulz, H. Zimdahl, H. Himmelbauer, H. Lehrach, H. J. Jacob, S. Bromberg, J. Gullings-Handley, M. I. Jensen-Seaman, A. E. Kwitek, J. Lazar, D. Pasko, P. J. Tonellato, S. Twigger, C. P. Ponting, J. M. Duarte, S. Rice, L. Goodstadt, S. A. Beatson, R. D. Emes, E. E. Winter, C. Webber, P. Brandt, G. Nyakatura, M. Adetobi, F. Chiaromonte, L. Elnitski, P. Eswara, R. C. Hardison, M. Hou, D. Kolbe, K. Makova, W. Miller, A. Nekrutenko, C. Riemer, S. Schwartz, J. Taylor, S. Yang, Y. Zhang, K. Lindpaintner, T. D. Andrews, M. Caccamo, M. Clamp, L. Clarke, V. Curwen, R. Durbin, E. Eyras, S. M. Searle, G. M. Cooper, S. Batzoglou, M. Brudno, A. Sidow, E. A. Stone, J. C. Venter, B. A. Payseur, G. Bourque, C. Lopez-Otin, X. S. Puente, K. Chakrabarti, S. Chatterji, C. Dewey, L. Pachter, N. Bray, V. B. Yap, A. Caspi, G. Tesler, P. A. Pevzner, D. Haussler, K. M. Roskin, R. Baertsch, H. Clawson, T. S. Furey, A. S. Hinrichs, D. Karolchik, W. J. Kent, K. R. Rosenbloom, H. Trumbower, M. Weirauch, D. N. Cooper, P. D. Stenson, B. Ma, M. Brent, M. Arumugam, D. Shteynberg, R. R. Copley, M. S. Taylor, H. Riethman, U. Mudunuri, J. Peterson, M. Guyer, A. Felsenfeld, S. Old, S. Mockrin and F. Collins, "Genome sequence of the Brown Norway rat yields insights into mammalian evolution." Nature, 2004 **428**(6982): 493-521.

Gilbert, N., S. Lutz, T. A. Morrish and J. V. Moran, "Multiple fates of l1 retrotransposition intermediates in cultured human cells." Mol Cell Biol, 2005 **25**(17): 7780-7795.

Gilbert, N., S. Lutz-Prigge and J. V. Moran, "Genomic deletions created upon LINE-1 retrotransposition." Cell, 2002 **110**(3): 315-325.

Gilbert, W., "Why genes in pieces?" Nature, 1978 **271**(5645): 501-501.

Girard, A., R. Sachidanandam, G. J. Hannon and M. A. Carmell, "A germline-specific class of small RNAs binds mammalian Piwi proteins." Nature, 2006 **442**(7099): 199-202.

Goff, S. P., "Retrovirus restriction factors." Mol Cell, 2004 **16**(6): 849-859.

Goll, M. G. and T. H. Bestor, "Eukaryotic cytosine methyltransferases." Annual Review of Biochemistry, 2005 **74**: 481-514.

Goodier, J. L., L. E. Cheung and H. H. Kazazian, Jr., "MOV10 RNA helicase is a potent inhibitor of retrotransposition in cells." PLoS Genet, 2012 **8**(10): e1002941.

Goodier, J. L., L. E. Cheung and H. H. Kazazian, Jr., "Mapping the LINE1 ORF1 protein interactome reveals associated inhibitors of human retrotransposition." Nucleic Acids Res, 2013.

Goodier, J. L., P. K. Mandal, L. Zhang and H. H. Kazazian, Jr., "Discrete subcellular partitioning of human retrotransposon RNAs despite a common mechanism of genome insertion." Hum Mol Genet, 2010 **19**(9): 1712-1725.

Goodier, J. L., E. M. Ostertag, K. A. Engleka, M. C. Seleme and H. H. Kazazian, Jr., "A potential role for the nucleolus in L1 retrotransposition." Hum Mol Genet, 2004 **13**(10): 1041-1048.

Goodier, J. L., E. M. Ostertag and H. H. Kazazian, Jr., "Transduction of 3'-flanking sequences is common in L1 retrotransposition." Hum Mol Genet, 2000 **9**(4): 653-657.

Goodier, J. L., L. Zhang, M. R. Vetter and H. H. Kazazian, Jr., "LINE-1 ORF1 protein localizes in stress granules with other RNA-binding proteins, including components of RNA interference RNA-induced silencing complex." Mol Cell Biol, 2007 **27**(18): 6469-6483.

Gregersen, L. H., M. Schueler, M. Munschauer, G. Mastrobuoni, W. Chen, S. Kempa, C. Dieterich and M. Landthaler, "MOV10 Is a 5' to 3' RNA helicase contributing to UPF1 mRNA target degradation by translocation along 3' UTRs." Mol Cell, 2014 **54**(4): 573-585.

Grimaldi, G., J. Skowronski and M. F. Singer, "Defining the beginning and end of KpnI family segments." EMBO J, 1984 **3**(8): 1753-1759.

Gunawardane, L. S., K. Saito, K. M. Nishida, K. Miyoshi, Y. Kawamura, T. Nagami, H. Siomi and M. C. Siomi, "A Slicer-Mediated Mechanism for Repeat-Associated siRNA 5' End Formation in Drosophila." Science, 2007 **315**(5818): 1587-1590.

Han, J. S., S. T. Szak and J. D. Boeke, "Transcriptional disruption by the L1 retrotransposon and implications for mammalian transcriptomes." Nature, 2004 **429**(6989): 268-274.

Han, K., J. Lee, T. J. Meyer, P. Remedios, L. Goodwin and M. A. Batzer, "L1 recombination-associated deletions generate human genomic variation." Proc Natl Acad Sci U S A, 2008 **105**(49): 19366-19371.

Han, K., S. K. Sen, J. Wang, P. A. Callinan, J. Lee, R. Cordaux, P. Liang and M. A. Batzer, "Genomic rearrangements by LINE-1 insertion-mediated deletion in the human and chimpanzee lineages." Nucleic Acids Res, 2005 **33**(13): 4040-4052.

Hancks, D. C., A. D. Ewing, J. E. Chen, K. Tokunaga and H. H. Kazazian, "Exon-trapping mediated by the human retrotransposon SVA." Genome Research, 2009 **19**(11): 1983-1991.

Hancks, D. C., J. L. Goodier, P. K. Mandal, L. E. Cheung and H. H. Kazazian, "Retrotransposition of marked SVA elements by human L1s in cultured cells." Human Molecular Genetics, 2011 **20**(17): 3386-3400.

Hancks, D. C. and H. H. Kazazian, Jr., "Active human retrotransposons: variation and disease." Curr Opin Genet Dev, 2012 **22**(3): 191-203.

Hata, K. and Y. Sakaki, "Identification of critical CpG sites for repression of L1 transcription by DNA methylation." Gene, 1997 **189**(2): 227-234.

Hattori, M., S. Kuhara, O. Takenaka and Y. Sakaki, "L1 family of repetitive DNA sequences in primates may be derived from a sequence encoding a reverse transcriptase-related protein." Nature, 1986 **321**(6070): 625-628.

Helman, E., M. S. Lawrence, C. Stewart, C. Sougnez, G. Getz and M. Meyerson, "Somatic retrotransposition in human cancer revealed by whole-genome and exome sequencing." Genome Res, 2014 **24**(7): 1053-1063.

Higashimoto, K., T. Maeda, J. Okada, Y. Ohtsuka, K. Sasaki, A. Hirose, M. Nomiyama, T. Takayanagi, R. Fukuzawa, H. Yatsuki, K. Koide, K. Nishioka, K. Joh, Y. Watanabe, K.-i. Yoshiura and H. Soejima, "Homozygous deletion of DIS3L2 exon 9 due to non-allelic homologous recombination between LINE-1s in a Japanese patient with Perlman syndrome." Eur J Hum Genet, 2013 **21**(11): 1316-1319.

Hohjoh, H. and M. F. Singer, "Cytoplasmic ribonucleoprotein complexes containing human LINE-1 protein and RNA." EMBO J, 1996 **15**(3): 630-639.

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

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): 6034-6043.

Holmes, S. E., B. A. Dombroski, C. M. Krebs, C. D. Boehm and H. H. Kazazian, Jr., "A new retrotransposable human L1 element from the LRE2 locus on chromosome 1q produces a chimaeric insertion." Nat Genet, 1994 **7**(2): 143-148.

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): 19765-19768.

Horn, A. V., S. Klawitter, U. Held, A. Berger, A. A. Jaguva Vasudevan, A. Bock, H. Hofmann, K. M. Hanschmann, J. H. Trosemeier, E. Flory, R. A. Jabulowsky, J. S. Han, J. Lower, R. Lower, C. Munk and G. G. Schumann, "Human LINE-1 restriction by APOBEC3C is deaminase independent and mediated by an ORF1p interaction that affects LINE reverse transcriptase activity." Nucleic Acids Res, 2013.

Howell, R. and K. Usdin, "The ability to form intrastrand tetraplexes is an evolutionarily conserved feature of the 3' end of L1 retrotransposons." Mol Biol Evol, 1997 **14**(2): 144-155.

Hrecka, K., C. Hao, M. Gierszewska, S. K. Swanson, M. Kesik-Brodacka, S. Srivastava, L. Florens, M. P. Washburn and J. Skowronski, "Vpx relieves

inhibition of HIV-1 infection of macrophages mediated by the SAMHD1 protein." Nature, 2011 **474**(7353): 658-661.

Iskow, R. C., M. T. McCabe, R. E. Mills, S. Torene, W. S. Pittard, A. F. Neuwald, E. G. Van Meir, P. M. Vertino and S. E. Devine, "Natural mutagenesis of human genomes by endogenous retrotransposons." Cell, 2010 **141**(7): 1253-1261.

Jacobs, F. M., D. Greenberg, N. Nguyen, M. Haeussler, A. D. Ewing, S. Katzman, B. Paten, S. R. Salama and D. Haussler, "An evolutionary arms race between KRAB zinc-finger genes ZNF91/93 and SVA/L1 retrotransposons." Nature, 2014 **516**(7530): 242-245.

Janousek, V., R. C. Karn and C. M. Laukaitis, "The role of retrotransposons in gene family expansions: insights from the mouse Abp gene family." BMC Evol Biol, 2013 **13**: 107.

Januszyk, K., P. W. Li, V. Villareal, D. Branciforte, H. Wu, Y. Xie, J. Feigon, J. A. Loo, S. L. Martin and R. T. Clubb, "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): 24893-24904.

Kajikawa, M., T. Sugano, R. Sakurai and N. Okada, "Low dependency of retrotransposition on the ORF1 protein of the zebrafish LINE, ZfL2-1." Gene, 2012 **499**(1): 41-47.

Kano, H., I. Godoy, C. Courtney, M. R. Vetter, G. L. Gerton, E. M. Ostertag and H. H. Kazazian, Jr., "L1 retrotransposition occurs mainly in embryogenesis and creates somatic mosaicism." Genes Dev, 2009 **23**(11): 1303-1312.

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): e181.

Kavanagh, D., D. Spitzer, P. H. Kothari, A. Shaikh, M. K. Liszewski, A. Richards and J. P. Atkinson, "New roles for the major human 3'-5' exonuclease TREX1 in human disease." Cell Cycle, 2008 **7**(12): 1718-1725.

Kazazian, H. H., Jr., C. Wong, H. Youssoufian, A. F. Scott, D. G. Phillips and S. E. Antonarakis, "Haemophilia A resulting from de novo insertion of L1 sequences represents a novel mechanism for mutation in man." Nature, 1988 **332**(6160): 164-166.

Khazina, E., V. Truffault, R. Buttner, S. Schmidt, M. Coles and O. Weichenrieder, "Trimeric structure and flexibility of the L1ORF1 protein in human L1 retrotransposition." Nat Struct Mol Biol, 2011 **18**(9): 1006-1014.

Khazina, E. and O. Weichenrieder, "Non-LTR retrotransposons encode noncanonical RRM domains in their first open reading frame." Proceedings of the National Academy of Sciences, 2009 **106**(3): 731-736.

Kohrman, D. C., J. B. Harris and M. H. Meisler, "Mutation detection in the med and medJ alleles of the sodium channel Scn8a. Unusual splicing due to a minor class AT-AC intron." The Journal of biological chemistry, 1996 **271**(29): 17576-17581.

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): 3237-3248.

Kulpa, D. A. and J. V. Moran, "Cis-preferential LINE-1 reverse transcriptase activity in ribonucleoprotein particles." Nat Struct Mol Biol, 2006 **13**(7): 655-660.

Kuramochi-Miyagawa, S., T. Watanabe, K. Gotoh, Y. Totoki, A. Toyoda, M. Ikawa, N. Asada, K. Kojima, Y. Yamaguchi, T. W. Ijiri, K. Hata, E. Li, Y. Matsuda, T. Kimura, M. Okabe, Y. Sakaki, H. Sasaki and T. Nakano, "DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes." Genes Dev, 2008 **22**(7): 908-917.

Laguette, N., B. Sobhian, N. Casartelli, M. Ringeard, C. Chable-Bessia, E. Segeral, A. Yatim, S. Emiliani, O. Schwartz and M. Benkirane, "SAMHD1 is the dendritic- and myeloid-cell-specific HIV-1 restriction factor counteracted by Vpx." Nature, 2011 **474**(7353): 654-657.

Lander, E. S., L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh, R. Funke, D. Gage, K. Harris, A. Heaford, J. Howland, L. Kann, J. Lehoczky, R. LeVine, P. McEwan, K. McKernan, J. Meldrim, J. P. Mesirov, C. Miranda, W. Morris, J. Naylor, C. Raymond, M. Rosetti, R. Santos, A. Sheridan, C. Sougnez, N. Stange-Thomann, N. Stojanovic, A. Subramanian, D. Wyman, J. Rogers, J. Sulston, R. Ainscough, S. Beck, D. Bentley, J. Burton, C. Clee, N. Carter, A. Coulson, R. Deadman, P. Deloukas, A. Dunham, I. Dunham, R. Durbin, L. French, D. Grafham, S. Gregory, T. Hubbard, S. Humphray, A. Hunt, M. Jones, C. Lloyd, A. McMurray, L. Matthews, S. Mercer, S. Milne, J. C. Mullikin, A. Mungall, R. Plumb, M. Ross, R. Shownkeen, S. Sims, R. H. Waterston, R. K. Wilson, L. W. Hillier, J. D. McPherson, M. A. Marra, E. R. Mardis, L. A. Fulton, A. T. Chinwalla, K. H. Pepin, W. R. Gish, S. L. Chissoe, M. C. Wendl, K. D. Delehaunty, T. L. Miner, A. Delehaunty, J. B. Kramer, L. L. Cook, R. S. Fulton, D. L. Johnson, P. J. Minx, S. W. Clifton, T. Hawkins, E. Branscomb, P. Predki, P. Richardson, S. Wenning, T. Slezak, N. Doggett, J. F. Cheng, A. Olsen, S. Lucas, C. Elkin, E. Uberbacher, M. Frazier, R. A. Gibbs, D. M. Muzny, S. E. Scherer, J. B. Bouck, E. J. Sodergren, K. C. Worley, C. M. Rives, J. H. Gorrell, M. L. Metzker, S. L. Naylor, R. S. Kucherlapati, D. L. Nelson, G. M. Weinstock, Y. Sakaki, A. Fujiyama, M. Hattori, T. Yada, A. Toyoda, T. Itoh, C. Kawagoe, H. Watanabe, Y. Totoki, T. Taylor, J. Weissenbach, R. Heilig, W. Saurin, F. Artiguenave, P. Brottier, T. Bruls, E. Pelletier, C. Robert, P. Wincker, D. R. Smith, L. Doucette-Stamm, M. Rubenfield, K. Weinstock, H. M. Lee, J. Dubois, A. Rosenthal, M. Platzer, G. Nyakatura, S. Taudien, A. Rump, H. Yang, J. Yu, J. Wang, G. Huang, J. Gu, L. Hood, L. Rowen, A. Madan, S. Qin, R. W. Davis, N. A. Federspiel, A. P. Abola, M. J. Proctor, R. M. Myers, J. Schmutz, M. Dickson, J. Grimwood, D. R. Cox, M. V. Olson, R. Kaul, C. Raymond, N. Shimizu, K. Kawasaki, S. Minoshima, G. A. Evans, M. Athanasiou, R. Schultz, B. A. Roe, F. Chen, H. Pan, J. Ramser, H. Lehrach, R. Reinhardt, W. R. McCombie, M. de la Bastide, N. Dedhia, H. Blocker, K. Hornischer, G. Nordsiek, R. Agarwala, L. Aravind, J. A. Bailey, A. Bateman, S. Batzoglou, E. Birney, P. Bork, D. G. Brown, C. B. Burge, L. Cerutti,

H. C. Chen, D. Church, M. Clamp, R. R. Copley, T. Doerks, S. R. Eddy, E. E. Eichler, T. S. Furey, J. Galagan, J. G. Gilbert, C. Harmon, Y. Hayashizaki, D. Haussler, H. Hermjakob, K. Hokamp, W. Jang, L. S. Johnson, T. A. Jones, S. Kasif, A. Kaspryzk, S. Kennedy, W. J. Kent, P. Kitts, E. V. Koonin, I. Korf, D. Kulp, D. Lancet, T. M. Lowe, A. McLysaght, T. Mikkelsen, J. V. Moran, N. Mulder, V. J. Pollara, C. P. Ponting, G. Schuler, J. Schultz, G. Slater, A. F. Smit, E. Stupka, J. Szustakowski, D. Thierry-Mieg, J. Thierry-Mieg, L. Wagner, J. Wallis, R. Wheeler, A. Williams, Y. I. Wolf, K. H. Wolfe, S. P. Yang, R. F. Yeh, F. Collins, M. S. Guyer, J. Peterson, A. Felsenfeld, K. A. Wetterstrand, A. Patrinos, M. J. Morgan, P. de Jong, J. J. Catanese, K. Osoegawa, H. Shizuya, S. Choi and Y. J. Chen, "Initial sequencing and analysis of the human genome." Nature, 2001 **409**(6822): 860-921.

Lawrence, C. J., Q. Dong, M. L. Polacco, T. E. Seigfried and V. Brendel, "MaizeGDB, the community database for maize genetics and genomics." Nucleic acids research, 2004 **32**(Database issue): D393-397.

Lee, E., R. Iskow, L. Yang, O. Gokcumen, P. Haseley, L. J. Luquette, 3rd, J. G. Lohr, C. C. Harris, L. Ding, R. K. Wilson, D. A. Wheeler, R. A. Gibbs, R. Kucherlapati, C. Lee, P. V. Kharchenko and P. J. Park, "Landscape of somatic retrotransposition in human cancers." Science, 2012 **337**(6097): 967-971.

Lee, J., K. Han, T. J. Meyer, H.-S. Kim and M. A. Batzer, "Chromosomal Inversions between Human and Chimpanzee Lineages Caused by Retrotransposons." PLoS One, 2008 **3**(12): e4047.

Leibold, D. M., G. D. Swergold, M. F. Singer, R. E. Thayer, B. A. Dombroski and T. G. Fanning, "Translation of LINE-1 DNA elements in vitro and in human cells." Proc Natl Acad Sci U S A, 1990 **87**(18): 6990-6994.

Li, P. W., J. Li, S. L. Timmerman, L. A. Krushel and S. L. Martin, "The dicistronic RNA from the mouse LINE-1 retrotransposon contains an internal ribosome entry site upstream of each ORF: implications for retrotransposition." Nucleic Acids Res, 2006 **34**(3): 853-864.

Li, W., L. Prazak, N. Chatterjee, S. Gruninger, L. Krug, D. Theodorou and J. Dubnau, "Activation of transposable elements during aging and neuronal decline in Drosophila." Nat Neurosci, 2013 **16**(5): 529-531.

Li, X., J. Zhang, R. Jia, V. Cheng, X. Xu, W. Qiao, F. Guo, C. Liang and S. Cen, "The MOV10 helicase inhibits LINE-1 mobility." J Biol Chem, 2013 **288**(29): 21148-21160.

Lin, L., J. Faraco, R. Li, H. Kadotani, W. Rogers, X. Lin, X. Qiu, P. J. de Jong, S. Nishino and E. Mignot, "The sleep disorder canine narcolepsy is caused by a mutation in the hypocretin (orexin) receptor 2 gene." Cell, 1999 **98**(3): 365-376.

Long, M., E. Betran, K. Thornton and W. Wang, "The origin of new genes: glimpses from the young and old." Nat Rev Genet, 2003 **4**(11): 865-875.

Luan, D. D., M. H. Korman, J. L. Jakubczak and T. H. Eickbush, "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): 595-605.

Macfarlane, C. M., P. Collier, R. Rahbari, C. R. Beck, J. F. Wagstaff, S. Igoe, J. V. Moran and R. M. Badge, "Transduction-Specific ATLAS Reveals a Cohort of Highly Active L1 Retrotransposons in Human Populations." Human Mutation, 2013 **34**(7): 974-985.

Macia, A., M. Munoz-Lopez, J. L. Cortes, R. K. Hastings, S. Morell, G. Lucena-Aguilar, J. A. Marchal, R. M. Badge and J. L. Garcia-Perez, "Epigenetic control of retrotransposon expression in human embryonic stem cells." Mol Cell Biol, 2011 **31**(2): 300-316.

Malik, H. S., W. D. Burke and T. H. Eickbush, "The age and evolution of non-LTR retrotransposable elements." Mol Biol Evol, 1999 **16**(6): 793-805.

Malki, S., G. W. van der Heijden, K. A. O'Donnell, S. L. Martin and A. Bortvin, "A role for retrotransposon LINE-1 in fetal oocyte attrition in mice." Developmental cell, 2014 **29**(5): 521-533.

Marques, A. C., I. Dupanloup, N. Vinckenbosch, A. Reymond and H. Kaessmann, "Emergence of Young Human Genes after a Burst of Retroposition in Primates." PLoS Biol, 2005 **3**(11): e357.

Martienssen, R., "Transposons, DNA methylation and gene control." Trends in Genetics, 1998 **14**(7): 263-264.

Martin, S. L., "Ribonucleoprotein particles with LINE-1 RNA in mouse embryonal carcinoma cells." Mol Cell Biol, 1991 **11**(9): 4804-4807.

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): 5383-5392.

Martin, S. L., D. Branciforte, D. Keller and D. L. Bain, "Trimeric structure for an essential protein in L1 retrotransposition." Proc Natl Acad Sci U S A, 2003 **100**(24): 13815-13820.

Martin, S. L., D. Bushman, F. Wang, P. W. Li, A. Walker, J. Cummiskey, D. Branciforte and M. C. Williams, "A single amino acid substitution in ORF1 dramatically decreases L1 retrotransposition and provides insight into nucleic acid chaperone activity." Nucleic Acids Res, 2008 **36**(18): 5845-5854.

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): 467-475.

Martin, S. L., M. Cruceanu, D. Branciforte, P. Wai-Lun Li, S. C. Kwok, R. S. Hodges and M. C. Williams, "LINE-1 retrotransposition requires the nucleic acid chaperone activity of the ORF1 protein." J Mol Biol, 2005 **348**(3): 549-561.

Martin, S. L., J. Li and J. A. Weisz, "Deletion analysis defines distinct functional domains for protein-protein and nucleic acid interactions in the ORF1 protein of mouse LINE-1." J Mol Biol, 2000 **304**(1): 11-20.

Mathias, S. L., A. F. Scott, H. H. Kazazian, Jr., J. D. Boeke and A. Gabriel, "Reverse transcriptase encoded by a human transposable element." Science, 1991 **254**(5039): 1808-1810.

McCarrey, J. R., "Molecular evolution of the human Pgk-2 retroposon." Nucleic Acids Res, 1990 **18**(4): 949-955.

McCarrey, J. R. and K. Thomas, "Human testis-specific PGK gene lacks introns and possesses characteristics of a processed gene." Nature, 1987 **326**(6112): 501-505.

McClintock, B., "The origin and behavior of mutable loci in maize." Proc Natl Acad Sci U S A, 1950 **36**(6): 344-355.

McClintock, B., "Chromosome organization and genic expression." Cold Spring Harb Symp Quant Biol, 1951 **16**: 13-47.

Mcclintock, B., "Some Parallels between Gene Control Systems in Maize and in Bacteria." American Naturalist, 1961 **95**(884): 265-277.

McClintock, B., "The Significance of Responses of the Genome to Challenge." Science, 1984 **226**(4676): 792-801.

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): 11533-11537.

Medstrand, P. and D. L. Mager, "Human-Specific Integrations of the HERV-K Endogenous Retrovirus Family." Journal of Virology, 1998 **72**(12): 9782-9787.

Miki, Y., T. Katagiri, F. Kasumi, T. Yoshimoto and Y. Nakamura, "Mutation analysis in the BRCA2 gene in primary breast cancers." Nat Genet, 1996 **13**(2): 245-247.

Miki, Y., I. Nishisho, A. Horii, Y. Miyoshi, J. Utsunomiya, K. W. Kinzler, B. Vogelstein and Y. Nakamura, "Disruption of the APC Gene by a Retrotransposal Insertion of L1 Sequence in a Colon Cancer." Cancer Research, 1992 **52**(3): 643-645.

Mine, M., J. M. Chen, M. Brivet, I. Desguerre, D. Marchant, P. de Lonlay, A. Bernard, C. Ferec, M. Abitbol, D. Ricquier and C. Marsac, "A large genomic deletion in the PDHX gene caused by the retrotranspositional insertion of a full-length LINE-1 element." Hum Mutat, 2007 **28**(2): 137-142.

Mitra, R., X. Li, A. l. Kapusta, D. Mayhew, R. D. Mitra, C. d. Feschotte and N. L. Craig, "Functional characterization of piggyBat from the bat Myotis lucifugus unveils an active mammalian DNA transposon." Proceedings of the National Academy of Sciences, 2013 **110**(1): 234-239.

Moran, J. V., R. J. DeBerardinis and H. H. Kazazian, Jr., "Exon shuffling by L1 retrotransposition." Science, 1999 **283**(5407): 1530-1534.

Moran, J. V., S. E. Holmes, T. P. Naas, R. J. DeBerardinis, J. D. Boeke and H. H. Kazazian, Jr., "High frequency retrotransposition in cultured mammalian cells." Cell, 1996 **87**(5): 917-927.

Morse, B., P. G. Rotherg, V. J. South, J. M. Spandorfer and S. M. Astrin, "Insertional mutagenesis of the myc locus by a LINE-1 sequence in a human breast carcinoma." Nature, 1988 **333**(6168): 87-90.

Muckenfuss, H., M. Hamdorf, U. Held, M. Perkovic, J. Lower, K. Cichutek, E. Flory, G. G. Schumann and C. Munk, "APOBEC3 proteins inhibit human LINE-1 retrotransposition." J Biol Chem, 2006 **281**(31): 22161-22172.

Mulhardt, C., M. Fischer, P. Gass, D. Simon-Chazottes, J.-L. Guenet, J. Kuhse, H. Betz and C.-M. Becker, "The spastic mouse: Aberrant splicing of glycine receptor Œ≤ subunit mRNA caused by intronic insertion of Ll element." Neuron, 1994 **13**(4): 1003-1015.

Muotri, A. R., V. T. Chu, M. C. Marchetto, W. Deng, J. V. Moran and F. H. Gage, "Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition." Nature, 2005 **435**(7044): 903-910.

Muotri, A. R., M. C. Marchetto, N. G. Coufal, R. Oefner, G. Yeo, K. Nakashima and F. H. Gage, "L1 retrotransposition in neurons is modulated by MeCP2." Nature, 2010 **468**(7322): 443-446.

Nakamura, M., N. Okada and M. Kajikawa, "Self-interaction, nucleic acid binding, and nucleic acid chaperone activities are unexpectedly retained in the unique ORF1p of zebrafish LINE." Mol Cell Biol, 2012 **32**(2): 458-469.

Nisole, S. b., C. Lynch, J. P. Stoye and M. W. Yap, "A Trim5-cyclophilin A fusion protein found in owl monkey kidney cells can restrict HIV-1." Proceedings of the National Academy of Sciences of the United States of America, 2004 **101**(36): 13324-13328.

O'Donnell, K. A., W. An, C. T. Schrum, S. J. Wheelan and J. D. Boeke, "Controlled insertional mutagenesis using a LINE-1 (ORFeus) gene-trap mouse model." Proc Natl Acad Sci U S A, 2013 **110**(29): E2706-2713.

Ohno, S., "So much "junk" DNA in our genome." Brookhaven Symp Biol, 1972 **23**: 366-370.

Orgel, L. E. and F. H. C. Crick, "Selfish DNA: the ultimate parasite." Nature, 1980 **284**(5757): 604-607.

Ostertag, E. M., R. J. DeBerardinis, J. L. Goodier, Y. Zhang, N. Yang, G. L. Gerton and H. H. Kazazian, Jr., "A mouse model of human L1 retrotransposition." Nat Genet, 2002 **32**(4): 655-660.

Ostertag, E. M., J. L. Goodier, Y. Zhang and H. H. Kazazian, Jr., "SVA elements are nonautonomous retrotransposons that cause disease in humans." Am J Hum Genet, 2003 **73**(6): 1444-1451.

Pace, J. K. and C. d. Feschotte, "The evolutionary history of human DNA transposons: Evidence for intense activity in the primate lineage." Genome Research, 2007 **17**(4): 422-432.

Parker, H. G., B. M. VonHoldt, P. Quignon, E. H. Margulies, S. Shao, D. S. Mosher, T. C. Spady, A. Elkahloun, M. Cargill, P. G. Jones, C. L. Maslen, G. M. Acland, N. B. Sutter, K. Kuroki, C. D. Bustamante, R. K. Wayne and E. A. Ostrander, "An Expressed Fgf4 Retrogene Is Associated with Breed-Defining Chondrodysplasia in Domestic Dogs." Science, 2009 **325**(5943): 995-998.

Pei, B., C. Sisu, A. Frankish, C. d. Howald, L. Habegger, X. J. Mu, R. Harte, S. Balasubramanian, A. Tanzer, M. Diekhans, A. Reymond, T. J. Hubbard, J. Harrow and M. B. Gerstein, "The GENCODE pseudogene resource." Genome Biology, 2012 **13**(9): R51-R51.

Perepelitsa-Belancio, V. and P. Deininger, "RNA truncation by premature polyadenylation attenuates human mobile element activity." Nat Genet, 2003 **35**(4): 363-366.

Perrat, P. N., S. DasGupta, J. Wang, W. Theurkauf, Z. Weng, M. Rosbash and S. Waddell, "Transposition-driven genomic heterogeneity in the Drosophila brain." Science, 2013 **340**(6128): 91-95.

Perreault, J., J. F. Noel, F. Briere, B. Cousineau, J. F. Lucier, J. P. Perreault and G. Boire, "Retropseudogenes derived from the human Ro/SS-A autoantigen-associated hY RNAs." Nucleic Acids Res, 2005 **33**(6): 2032-2041.

Pezic, D., S. A. Manakov, R. Sachidanandam and A. A. Aravin, "piRNA pathway targets active LINE1 elements to establish the repressive H3K9me3 mark in germ cells." Genes Dev, 2014 **28**(13): 1410-1428.

Pickeral, O. K., W. Maka≈Çowski, M. S. Boguski and J. D. Boeke, "Frequent Human Genomic DNA Transduction Driven by LINE-1 Retrotransposition." Genome Research, 2000 **10**(4): 411-415.

Pink, R. C., K. Wicks, D. P. Caley, E. K. Punch, L. Jacobs and D. R. Carter, "Pseudogenes: pseudo-functional or key regulators in health and disease?" RNA, 2011 **17**(5): 792-798.

Piskareva, O., S. Denmukhametova and V. Schmatchenko, "Functional reverse transcriptase encoded by the human LINE-1 from baculovirus-infected insect cells." Protein Expr Purif, 2003 **28**(1): 125-130.

Piskareva, O., C. Ernst, N. Higgins and V. Schmatchenko, "The carboxy-terminal segment of the human LINE-1 ORF2 protein is involved in RNA binding." FEBS Open Bio, 2013 **3**(0): 433-437.

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): 661-668.

Prak, E. T., A. W. Dodson, E. A. Farkash and H. H. Kazazian, Jr., "Tracking an embryonic L1 retrotransposition event." Proc Natl Acad Sci U S A, 2003 **100**(4): 1832-1837.

Puszyk, W., T. Down, D. Grimwade, C. Chomienne, R. J. Oakey, E. Solomon and F. Guidez, "The epigenetic regulator PLZF represses L1 retrotransposition in germ and progenitor cells." EMBO J, 2013 **32**(13): 1941-1952.

Raiz, J., A. Damert, S. Chira, U. Held, S. Klawitter, M. Hamdorf, J. Lower, W. H. Stratling, R. Lower and G. G. Schumann, "The non-autonomous retrotransposon SVA is trans-mobilized by the human LINE-1 protein machinery." Nucleic Acids Res, 2012 **40**(4): 1666-1683.

Ray, D. A., H. J. T. Pagan, M. L. Thompson and R. D. Stevens, "Bats with hATs: Evidence for Recent DNA Transposon Activity in Genus Myotis." Molecular Biology and Evolution, 2007 **24**(3): 632-639.

Reineke, L. C. and R. E. Lloyd, "Diversion of stress granules and P-bodies during viral infection." Virology, 2013 **436**(2): 255-267.

Repanas, K., N. Zingler, L. E. Layer, G. G. Schumann, A. Perrakis and O. Weichenrieder, "Determinants for DNA target structure selectivity of the human LINE-1 retrotransposon endonuclease." Nucleic Acids Res, 2007 **35**(14): 4914-4926.

Richardson, S. R., A. J. Doucet, H. C. Kopera, J. B. Moldovan, J. L. Garcia-Perez and J. V. Moran, "The Influence of LINE-1 and SINE Retrotransposons on Mammalian Genomes." Microbiology Spectrum, 2015 **3**(2).

Richardson, S. R., I. Narvaiza, R. A. Planegger, M. D. Weitzman and J. V. Moran, "APOBEC3A deaminates transiently exposed single-strand DNA during LINE-1 retrotransposition." Elife, 2014 **3**: e02008.

Rozmahel, R., H. H. Heng, A. M. Duncan, X. M. Shi, J. M. Rommens and L. C. Tsui, "Amplification of CFTR exon 9 sequences to multiple locations in the human genome." Genomics, 1997 **45**(3): 554-561.

Saito, K. and M. C. Siomi, "Small RNA-mediated quiescence of transposable elements in animals." Dev Cell, 2010 **19**(5): 687-697.

Sambrook, J., "Adenovirus amazes at Cold Spring Harbor." Nature, 1977 **268**(5616): 101-104.

Sassaman, D. M., B. A. Dombroski, J. V. Moran, M. L. Kimberland, T. P. Naas, R. J. DeBerardinis, A. Gabriel, G. D. Swergold and H. H. Kazazian, Jr., "Many human L1 elements are capable of retrotransposition." Nat Genet, 1997 **16**(1): 37-43.

Sayah, D. M., E. Sokolskaja, L. Berthoux and J. Luban, "Cyclophilin A retrotransposition into TRIM5 explains owl monkey resistance to HIV-1." Nature, 2004 **430**(6999): 569-573.

Scott, A. F., B. J. Schmeckpeper, M. Abdelrazik, C. T. Comey, B. O'Hara, J. P. Rossiter, T. Cooley, P. Heath, K. D. Smith and L. Margolet, "Origin of the human

L1 elements: proposed progenitor genes deduced from a consensus DNA sequence." Genomics, 1987 **1**(2): 113-125.

Segal, Y., B. Peissel, A. Renieri, M. de Marchi, A. Ballabio, Y. Pei and J. Zhou, "LINE-1 elements at the sites of molecular rearrangements in Alport syndrome-diffuse leiomyomatosis." Am J Hum Genet, 1999 **64**(1): 62-69.

Selker, E. U., N. A. Tountas, S. H. Cross, B. S. Margolin, J. G. Murphy, A. P. Bird and M. Freitag, "The methylated component of the Neurospora crassa genome." Nature, 2003 **422**(6934): 893-897.

Sen, S. K., K. Han, J. Wang, J. Lee, H. Wang, P. A. Callinan, M. Dyer, R. Cordaux, P. Liang and M. A. Batzer, "Human genomic deletions mediated by recombination between Alu elements." Am J Hum Genet, 2006 **79**(1): 41-53.

Sheehy, A. M., N. C. Gaddis, J. D. Choi and M. H. Malim, "Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein." Nature, 2002 **418**(6898): 646-650.

Shen, S. H., J. L. Slightom and O. Smithies, "A history of the human fetal globin gene duplication." Cell, 1981 **26**(2 Pt 2): 191-203.

Shukla, R., K. R. Upton, M. Munoz-Lopez, D. J. Gerhardt, M. E. Fisher, T. Nguyen, P. M. Brennan, J. K. Baillie, A. Collino, S. Ghisletti, S. Sinha, F. Iannelli, E. Radaelli, A. Dos Santos, D. Rapoud, C. Guettier, D. Samuel, G. Natoli, P. Carninci, F. D. Ciccarelli, J. L. Garcia-Perez, J. Faivre and G. J. Faulkner, "Endogenous retrotransposition activates oncogenic pathways in hepatocellular carcinoma." Cell, 2013 **153**(1): 101-111.

Sinnett, D., C. Richer, J. M. Deragon and D. Labuda, "Alu RNA secondary structure consists of two independent 7 SL RNA-like folding units." Journal of Biological Chemistry, 1991 **266**(14): 8675-8678.

Siomi, M. C., K. Sato, D. Pezic and A. A. Aravin, "PIWI-interacting small RNAs: the vanguard of genome defence." Nat Rev Mol Cell Biol, 2011 **12**(4): 246-258.

Skowronski, J., T. G. Fanning and M. F. Singer, "Unit-length line-1 transcripts in human teratocarcinoma cells." Mol Cell Biol, 1988 **8**(4): 1385-1397.

Skowronski, J. and M. F. Singer, "Expression of a cytoplasmic LINE-1 transcript is regulated in a human teratocarcinoma cell line." Proceedings of the National Academy of Sciences, 1985 **82**(18): 6050-6054.

Soifer, H. S., A. Zaragoza, M. Peyvan, M. A. Behlke and J. J. Rossi, "A potential role for RNA interference in controlling the activity of the human LINE-1 retrotransposon." Nucleic Acids Res, 2005 **33**(3): 846-856.

Solyom, S., A. D. Ewing, E. P. Rahrmann, T. Doucet, H. H. Nelson, M. B. Burns, R. S. Harris, D. F. Sigmon, A. Casella, B. Erlanger, S. Wheelan, K. R. Upton, R. Shukla, G. J. Faulkner, D. A. Largaespada and H. H. Kazazian, Jr., "Extensive somatic L1 retrotransposition in colorectal tumors." Genome Res, 2012 **22**(12): 2328-2338.

Soper, S. F., G. W. van der Heijden, T. C. Hardiman, M. Goodheart, S. L. Martin, P. de Boer and A. Bortvin, "Mouse maelstrom, a component of nuage, is essential for spermatogenesis and transposon repression in meiosis." Dev Cell, 2008 **15**(2): 285-297.

Speek, M., "Antisense promoter of human L1 retrotransposon drives transcription of adjacent cellular genes." Mol Cell Biol, 2001 **21**(6): 1973-1985.

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): 16837-16841.

Stetson, D. B., J. S. Ko, T. Heidmann and R. Medzhitov, "Trex1 prevents cell-intrinsic initiation of autoimmunity." Cell, 2008 **134**(4): 587-598.

Strichman-Almashanu, L. Z., M. Bustin and D. Landsman, "Retroposed copies of the HMG genes: a window to genome dynamics." Genome Res, 2003 **13**(5): 800-812.

Suter, C. M., D. I. Martin and R. L. Ward, "Hypomethylation of L1 retrotransposons in colorectal cancer and adjacent normal tissue." Int J Colorectal Dis, 2004 **19**(2): 95-101.

Swergold, G. D., "Identification, characterization, and cell specificity of a human LINE-1 promoter." Mol Cell Biol, 1990 **10**(12): 6718-6729.

Symer, D. E., C. Connelly, S. T. Szak, E. M. Caputo, G. J. Cost, G. Parmigiani and J. D. Boeke, "Human l1 retrotransposition is associated with genetic instability in vivo." Cell, 2002 **110**(3): 327-338.

Taniguchi-Ikeda, M., K. Kobayashi, M. Kanagawa, C. C. Yu, K. Mori, T. Oda, A. Kuga, H. Kurahashi, H. O. Akman, S. DiMauro, R. Kaji, T. Yokota, S. Takeda and T. Toda, "Pathogenic exon-trapping by SVA retrotransposon and rescue in Fukuyama muscular dystrophy." Nature, 2011 **478**(7367): 127-131.

Taylor, M. S., J. Lacava, P. Mita, K. R. Molloy, C. R. Huang, D. Li, E. M. Adney, H. Jiang, K. H. Burns, B. T. Chait, M. P. Rout, J. D. Boeke and L. Dai, "Affinity Proteomics Reveals Human Host Factors Implicated in Discrete Stages of LINE-1 Retrotransposition." Cell, 2013 **155**(5): 1034-1048.

Tchenio, T., J. F. Casella and T. Heidmann, "Members of the SRY family regulate the human LINE retrotransposons." Nucleic Acids Res, 2000 **28**(2): 411-415.

Thayer, R. E., M. F. Singer and T. G. Fanning, "Undermethylation of specific LINE-1 sequences in human cells producing a LINE-1-encoded protein." Gene, 1993 **133**(2): 273-277.

Thompson, C. B., "New insights into V(D)J recombination and its role in the evolution of the immune system." Immunity, 1995 **3**(5): 531-539.

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): 1520-1524.

Tubio, J. M., Y. Li, Y. S. Ju, I. Martincorena, S. L. Cooke, M. Tojo, G. Gundem, C. P. Pipinikas, J. Zamora, K. Raine, A. Menzies, P. Roman-Garcia, A. Fullam, M. Gerstung, A. Shlien, P. S. Tarpey, E. Papaemmanuil, S. Knappskog, P. Van Loo, M. Ramakrishna, H. R. Davies, J. Marshall, D. C. Wedge, J. W. Teague, A. P. Butler, S. Nik-Zainal, L. Alexandrov, S. Behjati, L. R. Yates, N. Bolli, L. Mudie, C. Hardy, S. Martin, S. McLaren, S. O'Meara, E. Anderson, M. Maddison, S. Gamble, I. B. C. Group, I. B. C. Group, I. P. C. Group, C. Foster, A. Y. Warren, H. Whitaker, D. Brewer, R. Eeles, C. Cooper, D. Neal, A. G. Lynch, T. Visakorpi, W. B. Isaacs, L. van't Veer, C. Caldas, C. Desmedt, C. Sotiriou, S. Aparicio, J. A. Foekens, J. E. Eyfjord, S. R. Lakhani, G. Thomas, O. Myklebost, P. N. Span, A. L. Borresen-Dale, A. L. Richardson, M. Van de Vijver, A. Vincent-Salomon, G. G. Van den Eynden, A. M. Flanagan, P. A. Futreal, S. M. Janes, G. S. Bova, M. R. Stratton, U. McDermott and P. J. Campbell, "Mobile DNA in cancer. Extensive transduction of nonrepetitive DNA mediated by L1 retrotransposition in cancer genomes." Science, 2014 **345**(6196): 1251343.

Ullu, E. and C. Tschudi, "Alu sequences are processed 7SL RNA genes." Nature, 1984 **312**(5990): 171-172.

Upton, K. R., D. J. Gerhardt, J. S. Jesuadian, S. R. Richardson, F. J. Sanchez-Luque, G. O. Bodea, A. D. Ewing, C. Salvador-Palomeque, M. S. van der Knaap, P. M. Brennan, A. Vanderver and G. J. Faulkner, "Ubiquitous l1 mosaicism in hippocampal neurons." Cell, 2015 **161**(2): 228-239.

Vagin, V. V., A. Sigova, C. Li, H. Seitz, V. Gvozdev and P. D. Zamore, "A distinct small RNA pathway silences selfish genetic elements in the germline." Science, 2006 **313**(5785): 320-324.

van den Hurk, J. A., I. C. Meij, M. C. Seleme, H. Kano, K. Nikopoulos, L. H. Hoefsloot, E. A. Sistermans, I. J. de Wijs, A. Mukhopadhyay, A. S. Plomp, P. T. de Jong, H. H. Kazazian and F. P. Cremers, "L1 retrotransposition can occur early in human embryonic development." Hum Mol Genet, 2007 **16**(13): 1587-1592.

van den Hurk, J. A., D. J. van de Pol, B. Wissinger, M. A. van Driel, L. H. Hoefsloot, I. J. de Wijs, L. I. van den Born, J. R. Heckenlively, H. G. Brunner, E. Zrenner, H. H. Ropers and F. P. Cremers, "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): 268-275.

Vinckenbosch, N., I. Dupanloup and H. Kaessmann, "Evolutionary fate of retroposed gene copies in the human genome." Proc Natl Acad Sci U S A, 2006 **103**(9): 3220-3225.

Virgen, C. A., Z. Kratovac, P. D. Bieniasz and T. Hatziioannou, "Independent genesis of chimeric TRIM5-cyclophilin proteins in two primate species." Proceedings of the National Academy of Sciences of the United States of America, 2008 **105**(9): 3563-3568.

Volkman, H. E. and D. B. Stetson, "The enemy within: endogenous retroelements and autoimmune disease." Nat Immunol, 2014 **15**(5): 415-422.

Wallace, N., B. J. Wagstaff, P. L. Deininger and A. M. Roy-Engel, "LINE-1 ORF1 protein enhances Alu SINE retrotransposition." Gene, 2008 **419**(1-2): 1-6.

Walsh, C. P., J. R. Chaillet and T. H. Bestor, "Transcription of IAP endogenous retroviruses is constrained by cytosine methylation." Nat Genet, 1998 **20**(2): 116-117.

Wang, H., J. Xing, D. Grover, D. J. Hedges, K. Han, J. A. Walker and M. A. Batzer, "SVA elements: a hominid-specific retroposon family." J Mol Biol, 2005 **354**(4): 994-1007.

Watanabe, T., Y. Totoki, A. Toyoda, M. Kaneda, S. Kuramochi-Miyagawa, Y. Obata, H. Chiba, Y. Kohara, T. Kono, T. Nakano, M. A. Surani, Y. Sakaki and H. Sasaki, "Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes." Nature, 2008 **453**(7194): 539-543.

Weber, M. J., "Mammalian Small Nucleolar RNAs Are Mobile Genetic Elements." PLoS Genet, 2006 **2**(12): e205.

Wei, W., N. Gilbert, S. L. Ooi, J. F. Lawler, E. M. Ostertag, H. H. Kazazian, J. D. Boeke and J. V. Moran, "Human L1 retrotransposition: cis preference versus trans complementation." Mol Cell Biol, 2001 **21**(4): 1429-1439.

Weichenrieder, O., K. Repanas and A. Perrakis, "Crystal structure of the targeting endonuclease of the human LINE-1 retrotransposon." Structure (Camb), 2004 **12**(6): 975-986.

Wilson, S. J., B. L. J. Webb, L. M. J. Ylinen, E. Verschoor, J. L. Heeney and G. J. Towers, "Independent evolution of an antiviral TRIMCyp in rhesus macaques." Proceedings of the National Academy of Sciences of the United States of America, 2008 **105**(9): 3557-3562.

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

Xing, J., H. Wang, V. P. Belancio, R. Cordaux, P. L. Deininger and M. A. Batzer, "Emergence of primate genes by retrotransposon-mediated sequence transduction." Proceedings of the National Academy of Sciences, 2006 **103**(47): 17608-17613.

Xiong, Y. and T. H. Eickbush, "Origin and evolution of retroelements based upon their reverse transcriptase sequences." EMBO J, 1990 **9**(10): 3353-3362.

Yajima, I., S. Sato, T. Kimura, K. Yasumoto, S. Shibahara, C. R. Goding and H. Yamamoto, "An L1 element intronic insertion in the black-eyed white (Mitf[mi-bw]) gene: the loss of a single Mitf isoform responsible for the pigmentary defect and inner ear deafness." Hum Mol Genet, 1999 **8**(8): 1431-1441.

Yan, N. and Z. J. Chen, "Intrinsic antiviral immunity." Nat Immunol, 2012 **13**(3): 214-222.

Yang, J., H. S. Malik and T. H. Eickbush, "Identification of the endonuclease domain encoded by R2 and other site-specific, non-long terminal repeat retrotransposable elements." Proc Natl Acad Sci U S A, 1999 **96**(14): 7847-7852.

Yang, N. and H. H. Kazazian, Jr., "L1 retrotransposition is suppressed by endogenously encoded small interfering RNAs in human cultured cells." Nat Struct Mol Biol, 2006 **13**(9): 763-771.

Yang, N., L. Zhang, Y. Zhang and H. H. Kazazian, Jr., "An important role for RUNX3 in human L1 transcription and retrotransposition." Nucleic Acids Res, 2003 **31**(16): 4929-4940.

Yoder, J. A., C. P. Walsh and T. H. Bestor, "Cytosine methylation and the ecology of intragenomic parasites." Trends Genet, 1997 **13**(8): 335-340.

Yohn, C. T., Z. Jiang, S. D. McGrath, K. E. Hayden, P. Khaitovich, M. E. Johnson, M. Y. Eichler, J. D. McPherson, S. Zhao, S. Paabo and E. E. Eichler, "Lineage-specific expansions of retroviral insertions within the genomes of African great apes but not humans and orangutans." PLoS biology, 2005 **3**(4): e110.

Yu, F., N. Zingler, G. Schumann and W. H. Stratling, "Methyl-CpG-binding protein 2 represses LINE-1 expression and retrotransposition but not Alu transcription." Nucleic Acids Res, 2001 **29**(21): 4493-4501.

Zhang, A., B. Dong, A. J. Doucet, J. B. Moldovan, J. V. Moran and R. H. Silverman, "RNase L restricts the mobility of engineered retrotransposons in cultured human cells." Nucleic Acids Res, 2014 **42**(6): 3803-3820.

Zhang, Z., P. M. Harrison, Y. Liu and M. Gerstein, "Millions of years of evolution preserved: a comprehensive catalog of the processed pseudogenes in the human genome." Genome Res, 2003 **13**(12): 2541-2558.

Zhao, K., J. Du, X. Han, J. L. Goodier, P. Li, X. Zhou, W. Wei, S. L. Evans, L. Li, W. Zhang, L. E. Cheung, G. Wang, H. H. Kazazian, Jr. and X. F. Yu, "Modulation of LINE-1 and Alu/SVA retrotransposition by Aicardi-Goutieres syndrome-related SAMHD1." Cell Rep, 2013 **4**(6): 1108-1115.

Zheng, K., J. Xiol, M. Reuter, S. Eckardt, N. A. Leu, K. J. McLaughlin, A. Stark, R. Sachidanandam, R. S. Pillai and P. J. Wang, "Mouse MOV10L1 associates with Piwi proteins and is an essential component of the Piwi-interacting RNA (piRNA) pathway." Proceedings of the National Academy of Sciences, 2010 **107**(26): 11841-11846.

Zheng, Y. H., K. T. Jeang and K. Tokunaga, "Host restriction factors in retroviral infection: promises in virus-host interaction." Retrovirology, 2012 **9**: 112.

Chapter 2

Mutations in Human ORF1p Disrupt The Function of LINE-1 Ribonucleoprotein Particles

It should be noted that some of the data images used in Figure 2.2 of this chapter were previously published in *PLoS Genetics* (Doucet et al. 2010) (see Figure 2.2 for details) and are presented here in strict accordance with PLoS journal policies and the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/). I performed all the experiments pertaining to Figure 2.2. I designed and carried out all experiments pertaining to Figures 2.3 and 2.4. The data depicted in Figures 2.3 and 2.4 have not been published.

Abstract

LINE-1 or L1 is the only active autonomous transposable element in the human genome. L1 mobilizes by the process of retrotransposition, which requires the activity of the two L1-encoded proteins (*i.e.*, ORF1p and ORF2p). ORF1p is an ~40 kDa nucleic acid binding protein and ORF2p encodes an ~150 kDa protein with endonuclease and reverse transcriptase activities. During L1 retrotransposition ORF1p and ORF2p bind to L1 RNA to form an L1 ribonucleoprotein particle (RNP), which is a required L1 retrotransposition intermediate. To elucidate the function of L1 ORF1p, we have used a genetic cell culture-based retrotransposition assay to monitor the effect of mutations in ORF1p on L1 retrotransposition. We also have used biochemical assays to determine how ORF1p mutations affect the formation and function of L1 RNPs. We report that mutations in the ORF1p coiled-coil (CC), RNA recognition motif

(RRM), and/or carboxyl-terminal (CTD) domains are incompatible with L1 retrotransposition. Biochemical analysis of these ORF1p mutants demonstrate that mutations that affect ORF1p RNA binding and/or mutations within the ORF1 CC domain impair the formation of L1 RNPs. Functional analyses of ORF1p mutant RNPs using an enzymatic assay to detect ORF2p RT activity indicate that mutations in ORF1p may result in reverse transcription of L1 RNA beginning within the 3' end of the L1. In sum, these data suggest that ORF1p is necessary for the formation of functional L1 RNPs.

Introduction

Long Interspersed Nuclear Element-1 (LINE-1, or L1) non-LTR retrotransposon sequences comprise ~17% of human DNA (Lander et al. 2001). L1 mobilizes throughout the genome by a process termed retrotransposition (Richardson et al. 2015). Although most genomic L1 sequences are inactive due to 5' truncations (Grimaldi et al. 1984, Lander et al. 2001) and various mutations, it is estimated that each cell contains at least 80-100 full length L1 sequences that are capable of retrotransposition (Sassaman et al. 1997, Brouha et al. 2003). Active L1 retrotransposition is known to cause genetic diseases (Hancks and Kazazian 2012) such as hemophilia A (Kazazian et al. 1988) and Duchenne muscular dystrophy (Narita et al. 1993). L1 retrotransposition also serves as an ongoing source of genetic sequence diversity that contributes to intra- and inter-individual variation in the human population (Cordaux and Batzer 2009, Beck et al. 2010).

An active full-length L1 is ~6 kb in length and consists of a 5' UTR that functions to promote L1 transcription (Swergold 1990, Becker et al. 1993, Athanikar et al. 2004), two non-overlapping open reading frames (ORFs) separated by a 63 bp intergenic spacer, and a 3' UTR that ends in a variable length stretch of poly-adenosine residues (Scott et al. 1987, Dombroski et al. 1991). The first L1 ORF (ORF1) encodes ORF1p, a ~40 kDa nucleic acid binding protein (Martin 1991, Holmes et al. 1992, Hohjoh and Singer 1996, Hohjoh and

Singer 1997) with nucleic acid chaperone activity (Martin and Bushman 2001). The second L1 ORF (ORF2) encodes ORF2p, an ~150 kDa protein, which contains endonuclease (EN) (Feng et al. 1996) and reverse transcriptase (RT) (Mathias et al. 1991, Dombroski et al. 1994) activities. Both L1-encoded proteins are required for L1 retrotransposition (Feng et al. 1996, Moran et al. 1996).

During the L1 retrotransposition process, a genomic L1 is transcribed in the nucleus and the resulting L1 mRNA is exported to the cytoplasm for translation. L1 translation occurs by an unconventional cap-dependent mechanism that ensures translation of both L1 ORFs (Leibold et al. 1990, McMillan and Singer 1993, Alisch et al. 2006, Dmitriev et al. 2007). Following translation of the L1 mRNA, ORF1p and ORF2p bind back to their encoding L1 mRNA (a phenomenon known as *cis*-preference (Esnault et al. 2000, Wei et al. 2001)) resulting in the formation of an L1 ribonucleoprotein particle (L1 RNP), which is an important retrotransposition intermediate (Kulpa and Moran 2005, Kulpa and Moran 2006). The L1 RNP then enters the nucleus (Kubo et al. 2006, Xie et al. 2013) where a new L1 copy is inserted into genomic DNA by the process of target site primed reverse transcription (TPRT) (Luan et al. 1993, Feng et al. 1996).

L1 retrotransposition requires functional ORF1p (Moran et al. 1996). Recent ORF1p structural studies have identified three distinct ORF1p functional domains (Januszyk et al. 2007, Khazina and Weichenrieder 2009, Khazina et al. 2011, Callahan et al. 2012). The amino-terminal region of ORF1p is composed of a coiled-coil (CC) domain that is thought to mediate the polymerization of ORF1p into trimers, the basic ORF1p functional unit (Martin et al. 2003, Khazina and Weichenrieder 2009, Khazina et al. 2011). Notably, the CC domain contains a putative leucine zipper motif that when mutated disrupts retrotransposition (Holmes et al. 1992, Hohjoh and Singer 1996, Hulme 2007, Khazina and Weichenrieder 2009, Doucet et al. 2010, Khazina et al. 2011). The central region of ORF1p contains an RNA recognition motif (RRM), which interacts with the ORF1p carboxyl terminus domain (CTD) to mediate RNA binding (Khazina and Weichenrieder 2009, Khazina et al. 2011). In addition to RNA binding, ORF1p

also exhibits putative nucleic acid chaperone activity, which can mediate the dissociation and/or annealing of double strand DNA *in vitro* (Martin and Bushman 2001, Martin et al. 2005, Khazina and Weichenrieder 2009, Callahan et al. 2012). Genetic and biochemical experiments have demonstrated that mutations in the ORF1p RRM or CTD domains disrupt RNA binding activity and/or nucleic acid chaperone activity, and thus are incompatible with L1 retrotransposition (Martin et al. 2005, Hulme 2007, Khazina and Weichenrieder 2009, Doucet et al. 2010, Khazina et al. 2011).

Early studies demonstrated that L1 ORF1p binds to L1 RNA and forms cytoplasmic RNP complexes that represent important L1 retrotransposition intermediates (Martin 1991, Martin and Branciforte 1993, Hohjoh and Singer 1996, Hohjoh and Singer 1997, Kulpa and Moran 2005). The subsequent development of an *in vitro* assay that monitors ORF2p RT activity (*i.e.*, LINE-1 Element Amplification Protocol or LEAP) in L1 RNPs provided the first biochemical evidence that ORF2p is also a constituent of L1 RNPs (Kulpa and Moran 2006). Notably, recent studies have employed epitope/RNA-tagging strategies coupled with biochemical and immunofluorescence microscopy methods to confirm that ORF1p, ORF2p, and L1 RNA constitute an L1 RNP (Doucet et al. 2010, Goodier et al. 2010, Taylor et al. 2013). Herein, we have expanded on these studies by analyzing how mutations in ORF1p impact the formation and function of L1 RNPs. We show that certain mutations in the ORF1p RRM and CTD perturb the normal biogenesis of L1 RNPs and are incompatible with retrotransposition. Analysis of ORF1p mutant RNPs suggest that mutations in ORF1p may affect L1 RNP function by modulating ORF2p RT activity.

Results

Mutations in ORF1p are incompatible with L1 retrotransposition

To analyze ORF1p function, we tested the effect of mutations in ORF1p on L1 retrotransposition using a cultured cell retrotransposition assay (Moran et

al. 1996, Wei et al. 2000). Briefly, HeLa cells were transfected with epitope-tagged L1 constructs (Figure 2.1) that express a T7 *gene10* epitope tag on the carboxyl terminus of ORF1p and also contain a neomycin retrotransposition indicator cassette (*mneoI*) within the L1 3' UTR (Figure 2.1) (Kulpa and Moran 2005). The *mneoI* cassette contains an antisense copy of the neomycin phosphotransferase gene, which is interrupted by an intron in the same transcriptional orientation as L1. This arrangement ensures that the neomycin gene is expressed only when the L1 transcript is spliced, reverse transcribed, and inserted into genomic DNA (Freeman et al. 1994, Moran et al. 1996). The resulting neomycin-resistant foci then provide a visual, quantitative readout of retrotransposition activity (Moran et al. 1996, Wei et al. 2000). The following L1 mutants were used to test the effects of point mutations in specific ORF1p domains (Figure 2.1): pDK103 ($REKG_{235\text{-}238}AAAA$); pDK105 ($RR_{261\text{-}262}AA$); pDK106 ($RR_{261\text{-}262}KK$); pDK108 ($R_{262}K$); pDK109 ($R_{261}K$); pDK116 ($YPAKLS_{282\text{-}287}AAAALA$); pLZC ($L_{93,\ 100,\ 107,\ 114}V$) (Hulme 2007); and pLZC/AA ($L_{93,\ 100,\ 107,\ 114}V$ + $RR_{261\text{-}262}AA$) (Hulme 2007). In agreement with previous reports (Moran et al. 1996, Kulpa and Moran 2005, Doucet et al. 2010), retrotransposition assays showed that pDK103, pDK105, pDK108, pDK106, and pDK116 were incapable of retrotransposition whereas, pDK108 retrotransposed at ~80% of wild type (pDK101) levels. The pLZC and pLZC/AA were previously demonstrated to abolish L1 retrotransposition activity (Hulme 2007, Doucet et al. 2010). Thus, the data suggest that the ORF1p CC, RRM, and CTD domains are required for L1 retrotransposition.

The Impact Of ORF1p Mutations On The Function Of L1 RNPs

We next tested the effect of ORF1p mutations on the assembly and function of L1 RNPs. Briefly, HeLa cells were transfected with wild type and mutant L1 constructs (Figure 2.1) and then L1 RNPs were isolated from transfected cell lysates using ultracentrifugation over sucrose cushions (Kulpa and Moran 2005, Kulpa and Moran 2006). Following ultracentrifugation, cellular RNP pellets were collected and then assayed for the presence of T7-tagged

ORF1p and L1 RNA. The presence of T7-tagged ORF1p and L1 RNA in the RNP pellet serves as an indication of L1 RNP formation. RNP fractions defined above also were assayed for L1 ORF2p reverse transcriptase activity using the LEAP assay (Kulpa and Moran 2006), which provides a readout for the presence of ORF2p in L1 RNPs and serves as an indication of L1 RNP function (*i.e.*, the ability to reverse transcribe L1 RNA).

In agreement with previous reports (Kulpa and Moran 2005, Kulpa and Moran 2006), RNPs derived from HeLa cells transfected with the wild type pDK101 contained a substantial enrichment of L1 ORF1p compared to untransfected HeLa cells (Figure 2.2B), thus confirming the presence of L1 RNPs in the crude RNP pellet fraction. In pDK101 LEAP reactions, a diffuse band between ~220 and ~400 bp was visualized on agarose gels indicating the presence of ORF2p reverse transcriptase activity (Figure 2.2C). Notably, control experiments showed that LEAP activity was absent from RNPs derived from pDK135 transfected cells (Figure 2.2C) (pDK135 contains a mutation in the ORF2 reverse transcriptase domain rendering the element unable to retrotranspose) (Moran et al. 1996, Wei et al. 2001, Kulpa and Moran 2006). Sequencing of pDK101 LEAP products revealed that L1 reverse transcription initiated at variable sites within the L1 poly-adenosine sequence, which accounts for LEAP product size variation (data not shown, and (Kulpa and Moran 2006)). Control M-MLV RT PCR reactions confirmed the presence of L1 RNA in pDK101 RNPs (Figure 2.2C).

RNA binding activity mutants (pDK103 ($REKG_{235\text{-}238}AAAA$); pDK105 ($RR_{261\text{-}262}AA$); pDK116 ($YPAKLS_{282\text{-}287}AAAALA$)): In agreement with previous reports (Kulpa and Moran 2005, Kulpa and Moran 2006), RNPs derived from HeLa cells transfected with pDK105, pDK116, and pDK103 contained less ORF1p than did wild type pDK101 control RNPs (Figure 2.2B). Similarly, whole cell lysates harvested from cells transfected with pDK105, pDK116, and pDK103 contained less ORF1p than lysates from pDK101 transfected cells (Figure 2.2B). Notably, the $R_{261}A$ mutation in pDK105 has been demonstrated to impair human ORF1p RNA binding activity *in vitro* (Khazina et al. 2011) and introduction of

($RR_{261-262}AA$) into the homologous region of mouse ORF1p has been shown to reduce the RNA binding activity and the stability of mouse ORF1p *in vitro* (Martin et al. 2005). These data suggest that mutations in the ORF1p RRM and CTD interfere with the ability of ORF1p to localize to RNPs and/or affect ORF1p stability.

In LEAP reactions using pDK105 RNPs, LEAP activity was marked by a diffuse band and the appearance of distinct lower bands (below 300bp) on LEAP gels (Figure 2.2C). Notably, pDK116 and pDK103 LEAP products generally appeared more diffuse than wild type pDK101 LEAP activity. In addition, pDK116 and pDK103 LEAP reactions often contained lower bands in LEAP gels similar to the lower bands observed in pDK105 LEAP gels (Figure 2.2C). Sequencing of pDK105 and pDK116 LEAP products indicated frequent initiation of L1 reverse transcription from within the 3' end of the L1 RNA (data not shown and (Kulpa and Moran 2006, Doucet et al. 2010)). Similar LEAP data also was reported for an L1 mutant that lacks the entire ORF1 sequence (Doucet et al. 2010). These data suggest that ORF2p can associate with L1 mRNA and carry out reverse transcription of L1 mRNA independently of ORF1p; and that ORF1p likely promotes the initiation of L1 reverse transcription from the L1 poly-adenosine tail (Kulpa and Moran 2006, Doucet et al. 2010).

Chaperone activity mutants (pDK106 ($RR_{261-262}KK$); pDK109 ($R_{261}K$)): Mutations in pDK106 and pDK109, which have previously been shown to affect ORF1p nucleic acid chaperone activity (Martin et al. 2005), did not affect the localization of ORF1p to RNPs (Figure 2.2B) and (Kulpa and Moran 2005). LEAP activity was similarly unaffected in pDK106 and pDK109 RNPs (Figure 2.2C). We did however notice the occasional appearance of lower molecular weight (less than ~300 bp) LEAP products in pDK106 and pDK109 LEAP reactions that were not observed in pDK101 LEAP reactions (Figure 2.2C). These data therefore suggest that although ORF1p may not be strictly required for L1 reverse transcriptase activity in RNPs, that certain ORF1p mutations affect the array of LEAP products observed in these assays.

Leucine zipper mutants (pLZC ($L_{93, 100, 107, 114}V$); pLZC/AA ($L_{93, 100, 107, 114}V$ + $RR_{261-262}AA$)): RNPs derived from pLZC ($L_{93, 100, 107, 114}V$) mutant transfected cells contained less ORF1p than pDK101 control RNPs (Figure 2.2B). LEAP activity from pLZC RNPs appeared to be less compared to pDK101 controls upon visual inspection of LEAP gels (Figure 2.2C). Quantitative LEAP experiments confirmed that pLZC RNP LEAP activity is decreased compared to wild type controls (Hulme 2007, Doucet et al. 2010). Notably, RNPs derived from mutant L1s that contain the $L_{93, 100, 107, 114}V$ mutations in ORF1p also contain less ORF2p than corresponding wild-type controls (Doucet et al. 2010). These data suggest that the LZC domain is required for the localization of ORF1p to L1 RNPs and that the $L_{93, 100, 107, 114}V$ mutations could impair either the localization of ORF2p to L1 RNPs and/or ORF2p RT activity.

Because the $L_{93, 100, 107, 114}V$ mutation in ORF1 appeared to exert a negative effect on the localization of ORF2p to L1 RNPs and/or ORF2p RT activity, we hypothesized that ORF1p might interact directly with ORF2p. To test this, we introduced the $RR_{261-262}AA$ mutation into the pLZC mutant construct to make an L1 mutant construct (pLZC/AA) that contains both the $L_{93, 100, 107, 114}V$ and the $RR_{261-262}AA$ mutation. We did so reasoning that since the $RR_{261-262}AA$ mutation impaired the localization of ORF1p to L1 RNPs without affecting the localization of ORF2p to RNPs, that the introduction of the $RR_{261-262}AA$ mutation into pLZC might relieve any negative effect that the $L_{93, 100, 107, 114}V$ ORF1 mutation could have on the localization of ORF2p to L1 RNPs and/or ORF2p RT activity. In L1 RNPs collected from pLZC/AA transfected cells, western blot experiments revealed a loss of ORF1p from pLZC/AA mutant RNPs compared to pDK101 wild type control RNPs (Figure 2.2B). LEAP activity also was decreased in pLZC/AA RNPs compared to pDK101 controls (Figure 2.2C), which has been verified by quantitative LEAP experiments using similar constructs (Hulme 2007, Doucet et al. 2010). Western blotting further revealed that pLZC/AA RNPs contained less ORF2p than corresponding wild-type controls (Doucet et al. 2010). Notably, LEAP activity from pLZC/AA RNPs exhibited lower bands in agarose gels similar to those observed in pDK105 LEAP gels (Figure 2.2C).

Thus, pLZC/AA mutant RNPs display phenotypes similar to both the pLZC and the pDK105 mutant RNPs. In sum, these data suggest that the ORF1p putative leucine zipper domain is required for the proper localization of ORF1p and ORF2p to L1 RNPs.

Engineering an ORF1p/ORF2p fusion protein to study the ORF1 LZC mutation

The above data suggest that mutations in ORF1p may negatively affect the localization of ORF1p and/or ORF2p to the L1 RNP. I further hypothesized that mutations in ORF1p also could affect ORF2p function (*e.g.,* ORF2p reverse transcriptase activity). Because some of the ORF1p mutations (*e.g.*, $RR_{261\text{-}262}AA$, $L_{93,\ 100,\ 107,\ 114}V$) affected the localization of ORF1p and/or ORF2p to L1 RNPs, an ORF1p-ORF2p fusion protein was constructed to facilitate an association between ORF1p and ORF2p so that I could test whether ORF1p mutations affected ORF2p function. The ORF1p-ORF2p fusion protein was constructed by substituting the ORF1 stop codon and the 63 bp intervening sequence between ORF1 and ORF2 with a single lysine residue, which effectively tethers ORF2p to the carboxyl terminus of ORF1p (Alisch et al. 2006). The ORF1p-ORF2p fusion protein construct (p1K2_TAP) was further modified to contain a TAP epitope tag on to the carboxyl terminus of ORF2 to facilitate the detection of the fusion protein using anti-TAP antibodies (Figure 2.3A).

First, to determine if the ORF1p-ORF2p fusion protein was expressed we transfected HeLa cells with p1K2_TAP and analyzed p1K2_TAP RNPs by western blot using anti-TAP antibodies. A prominent band of the expected size of an ORF1p-ORF2p fusion peptide (~190 kDa) was detected in p1K2_TAP RNPs using anti-TAP antibodies (Figure 2.3B: top panel). Notably, the ~190 kDa band was not present in RNPs derived from untransfected HeLa cells or from HeLa cells transfected with a wild-type L1 control (pAD2TE1) (Doucet et al. 2010), or an L1 that lacks the ORF1 sequence (pAD500) (Doucet et al. 2010) (Figure 2.3B: top panel). Several smaller molecular weight bands were detected in p1K2_TAP RNPs with the anti-TAP antibody (Figure 2.3B: top panel) suggesting that the

fusion could be processed into smaller peptides, or that the fusion had partially degraded during the sample preparation process (Figure 2.3B).

Next, we used an antibody specific to amino acids 31-49 of ORF1p to detect the amino terminal ORF1p end of the fusion peptide. Similar to blots probed with the anti-TAP antibody, we detected an ~190 kDa band in p1K2_TAP RNPs using the anti-ORF1p antibody (Figure 2.3B: bottom panel). The ~190 kDa band was noticeably absent from untransfected HeLa, pAD2TE1, and pAD500 RNPs (Figure 2.3B: bottom panel). A band corresponding to free ORF1p (~40 kDa) was detected in pAD2TE1 RNPs using the anti-ORF1p antibody (Figure 2.3B: bottom panel). In contrast, free ORF1p was not detected in p1K2_TAP RNPs (Figure 2.3B: bottom panel). Thus, free ORF1p is not expressed from p1K2_TAP and the ORF1p-ORF2p fusion peptide is not processed to produce free ORF1p.

In additional control experiments we engineered a mutant fusion peptide in which the nucleotide sequence corresponding to the first 255 amino acids were deleted from ORF1p (p1K2(RF1)) (Figure 2.3C), which would be expected to yield an ~160 kDa fusion peptide. Indeed, an ~160 kDa band was detected in RNPs derived from HeLa cells transfected with p1K2(RF1) (Figure 2.3C). We also introduced a mutation into the ORF1p sequence that inserted a premature stop codon into ORF1 of p1K2_TAP (p1K2(S119X)), which is expected to result in early termination of the fusion protein (Moran et al. 1996, Wei et al. 2001, Alisch et al. 2006). Western blots demonstrated a substantial decrease in expression of the fusion peptide from p1K2(S119X) compared to p1K2_TAP (Figure 2.3C). Notably, free ORF2p was not detected in the p1K2(S119X) lane suggesting that ORF2p translation does not initiate from internal sequences within the ORF1p-ORF2p fusion protein RNA. Northern blotting experiments using a probe to the L1 5' UTR (5UTR99) demonstrated L1 RNA expression levels were similar between p1K2(S119X), p1K2_TAP and pAD2TE1 (Figure 2.3D). Thus, the S119X mutation in ORF1 does not affect the accumulation of L1 RNA, and confirms previous data suggesting that the S119X mutation does not affect expression of L1 RNA (Moran et al. 1996, Wei et al. 2001, Alisch et al.

2006). In sum, the above data suggest that an ORF1p-ORF2p fusion peptide is expressed from p1K2_TAP.

The impact of the LZC mutation on the function of the ORF1p/ORF2p fusion protein

Next, we tested the effect of mutation in ORF1p on the function of the ORF1p-ORF2p fusion peptide. Although the ORF1p-ORF2p fusion protein does not retrotranspose in *cis*, it has been demonstrated to drive the retrotransposition of a retrotransposition reporter plasmid in a genetic *trans*-complementation assay (Figure 2.4A) (Wei et al. 2001, Alisch et al. 2006)). The genetic *trans*-complementation assay monitors the ability of an L1 construct lacking the *mneoI* retrotransposition indicator cassette (*i.e.*, a driver L1) to *trans*-mobilize a reporter mRNA consisting of the L1 5′ UTR, ORF1, and the spliced *mneoI* indicator cassette (*i.e.*, ORF1*mneoI*) (Figure 2.4A). By taking advantage of the ORF1p-ORF2p fusion protein, we addressed whether mutations in ORF1p affected the ability of the ORF1p-ORF2p fusion protein to *trans*-mobilize RNA derived from ORF1mneoI.

To test the effect of ORF1p mutations on the function of the ORF1p-ORF2p fusion protein in *trans*-complementation assays, we made the following ORF1p-ORF2p fusion protein constructs (Figure 2.3A): (p1K2(AA); $RR_{261\text{-}262}AA$), (p1K2(LZC); $L_{93,\ 100,\ 107,\ 114}V$), (p1K2(LZC/AA); $L_{93,\ 100,\ 107,\ 114}V$ + $RR_{261\text{-}262}AA$), (p1K2(S119X); S119X), (p1K2(RF1); ORF1Δ1-255). In control western blot experiments we observed similar expression levels of the fusion peptide and of the smaller bands from each of the fusion constructs (Figure 2.3E). Thus, the above mutations in ORF1 do not affect the expression of the ORF1p-ORF2p fusion peptide.

To determine the affect of ORF1p mutations when tethered to ORF2p to drive retrotransposition of a retrotransposition reporter construct in *trans*, the ORF1p-ORF2p fusion constructs described above were co-transfected with the ORF1*mneoI* reporter construct. Similar to a previous report (Alisch et al. 2006), the wild type p1K2_TAP efficiently *trans*-complemented the retrotransposition of

the ORF1*mneoI* reporter construct (Figure 2.4B). Notably, the *trans*-complementation efficiency of p1K2(LZC) was ~65% of wild type control p1K2 levels (Figure 2.4B). Thus, the LZC mutation in ORF1 does not significantly impair the ability of the fusion peptide to *trans*-complement the ORF1*mneoI* reporter construct. In contrast to p1K2(LZC), the *trans*-complementation efficiency of the p1K2(AA) and p1K2(LZC/AA) constructs were only ~38% and ~19% respectively of control levels (Figure 2.4B). The *trans*-complementation efficiency of the mutant fusion construct p1K2(RF1) in which the first 255 amino acids were deleted from ORF1p also was significantly decreased (~50% of wild type p1K2 control levels). Notably, within the first 255 amino acids of ORF1p are several amino acids that have been demonstrated to be important for *in vitro* ORF1p RNA binding activity including R235, R210, R211, K133, K137, K140, and R141 (Martin et al. 2005, Khazina et al. 2011). Thus, mutations that affect ORF1p RNA binding activity negatively affect ORF1*mneoI* *trans*-complementation. In additional control experiments, co-transfection of ORF1*mneoI* with an empty pCEP4 vector or the p1K2(S119X) mutant, which contains a missense mutation in ORF1 that results in early termination of translation (Moran et al. 1996, Wei et al. 2001, Alisch et al. 2006), expectedly did not result in the production of G418 resistant colonies (Figure 2.4B). Notably, the *trans*-complementation efficiency of pAD2TE1 was only ~12% of p1K2_TAP levels (Figure 2.4B), which could be explained by the considerable disparity in protein expression between p1K2_TAP and pAD2TE1 (Figure 2.3E). Thus, the data suggest that the LZC mutation ($L_{93,\ 100,\ 107,\ 114}V$) in ORF1 by itself, is insufficient to impair the ability of the ORF1p-ORF2p fusion peptide to *trans*-complement RNA derived from the ORF1p*mneoI* reporter construct.

Discussion

The L1 RNP is a critical retrotransposition intermediate (Martin 1991, Martin and Branciforte 1993, Hohjoh and Singer 1996, Kulpa and Moran 2005, Kulpa and Moran 2006, Doucet et al. 2010). We have extended recent analyses

of L1 RNPs (Kulpa and Moran 2005, Kulpa and Moran 2006) by examining how mutation of ORF1p impacts the function and formation L1 RNPs. We report that certain mutations in ORF1p, which are incompatible with retrotransposition, can affect the formation and/or stability of L1 RNPs. We have categorized these mutations based on how they impact L1 RNPs (Figure 2.5).

ORF1p RNA binding mutants (pDK103, pDK105, pDK116): Mutations that affect the ability of ORF1p to bind RNA result in RNPs that lack ORF1p but contain ORF2p and L1 reverse transcriptase activity (Type I RNPS, Figure 2.5B). Mutant ORF1p failed to localize to RNPs isolated from cells transfected with pDK105, pDK103, and pDK116. pDK105 carries the RR261-262AA mutation which has been shown to significantly reduce the RNA binding capacity of mouse (Martin et al. 2005) and human ORF1p (Khazina et al. 2011). Moreover, this mutation has also been shown to reduce the stability of mouse ORF1p although it does not appear to affect overall structure of either mouse or human ORF1p (Martin et al. 2005, Khazina et al. 2011). The R235A mutation in pDK103 occurs in the RRM domain and has also been shown to reduce the RNA binding activity of human ORF1p *in vitro* (Khazina et al. 2011).

Although the effect on RNA binding activity of the YPKALS$_{282\text{-}287}$AAALAA mutation has not been directly tested, several lines of evidence support that this mutation impairs the RNA binding activity. First, the YPKALS$_{282\text{-}287}$AAALAA mutation severely compromises the ability of ORF1p to localize to RNPs; second, the constellation of pDK116 LEAP products bear similarity to pDK105 LEAP products (*i.e.,* initiation of reverse transcription from within the 3' end of the L1 mRNA) (Kulpa and Moran 2006, Doucet et al. 2010); and third, the YPKALS$_{282\text{-}287}$AAALAA mutation impairs L1 cytoplasmic foci formation (Doucet et al. 2010). Thus, ORF1p RNA binding activity is important for the formation and/or stability of L1 RNPs.

RNPs that were derived from ORF1p RNA binding mutants contain ORF2p (Doucet et al. 2010) and exhibit robust LEAP activity (Doucet et al. 2010) (Figure 2.2C). In agreement with previous data (Kulpa and Moran 2006, Doucet

et al. 2010, Kopera et al. 2011), this result suggests that L1 ORF2p can bind to L1 RNA and carry out L1 cDNA synthesis in the absence of ORF1p. Notably, L1 elements that carry the RR261-262AA mutation or that entirely lack the ORF1 sequence are able to retrotranspose (albeit at very low levels) in Chinese hamster ovary cells that are deficient in the non-homologous end joining DNA repair pathway (Kopera et al. 2011). Interestingly, sequencing of LEAP products from pDK105, pDK116 and an L1 that does not express ORF1p indicated that L1 reverse transcription frequently initiated from within the L1 mRNA 3' end in RNPs devoid of ORF1p (Kulpa and Moran 2006, Doucet et al. 2010). Thus, in agreement with previous reports (Kulpa and Moran 2006, Doucet et al. 2010) the data suggest that ORF1p may help facilitate the initiation of L1 reverse transcription from the L1 poly-adenosine tail.

ORF1 chaperone activity mutants (pDK106, pDK109): The second group of RNPs (Figure 2.5B; Type II RNPS) includes RNPs derived from L1s that carried the R261K mutation (i.e., pDK106 & pDK109). The R261K affects ORF1p nucleic acid chaperone activity *in vitro* (Martin and Bushman 2001, Martin et al. 2005) without affecting RNA binding activity and is incompatible with retrotransposition (Kulpa and Moran 2005, Martin et al. 2005, Martin et al. 2008, Khazina et al. 2011). We showed that pDK106 and pDK109 RNPs contained wild type levels of ORF1p and exhibited wild type LEAP activity. Notably, immunofluorescence experiments also demonstrated that L1 cytoplasmic foci formation is unaffected by the R261K mutation (Doucet et al. 2010). Thus, the R261K mutation does not affect formation of L1 RNPs, which suggests that this mutation affects some other aspect of L1 RNP function (*e.g.*, nucleic acid chaperone activity).

ORF1 leucine zipper mutants (pLZC and pLZC/AA): The third category of RNPs includes L1s with mutations in the putative LZC motif within the ORF1p CC domain (Figure 2.5B; type III RNPs). Compared to wild-type controls, LZC mutant RNPs contained less ORF1p and ORF2p (Figure 2.2B and (Hulme 2007, Doucet et al. 2010). Notably, LZC mutant RNPs also and exhibited a reduction in LEAP activity (Figure 2.2C and (Hulme 2007, Doucet et al. 2010). These results

suggested the possibilities that the LZC mutation in ORF1p affected the ability of ORF2p to localize to RNPs and/or to reverse transcribe L1 RNA. Notably, the ORF1p-ORF2p fusion peptide carrying the LZC defect was able to efficiently drive the retrotransposition of the ORF1*mneoI* reporter construct (Figure 2.4). The data therefore suggest that the LZC mutation in ORF1p does not impair ORF2p reverse transcriptase function because *trans*-complementation of RNA derived from the ORF1*mneoI* reporter is dependent on ORF2p reverse transcriptase activity (Wei et al. 2001, Alisch et al. 2006). Thus, the most parsimonious explanation of the data is that the ORF1p LZC mutation most likely impairs the localization of ORF1p and/or ORF2p to RNPs. These data suggest the possibility that a subtle interplay between the ORF1p and ORF2p proteins may exist, which if disturbed disrupts the formation of functional L1 RNPs.

A previous report presented genetic evidence suggesting that an engineered ORF1p-ORF2p fusion peptide could *trans*-mobilize ORF1*mneoI* (Alisch et al. 2006). In this study we have provided the first empirical evidence that a similar engineered ORF1p-ORF2p fusion peptide (p1K2_TAP) can be expressed in human cells and that it is capable of *trans*-mobilizing other cellular mRNAs (*i.e*., ORF1*mneoI*). We demonstrated that an ~190 kDa ORF1p-ORF2p fusion protein was expressed from p1K2_TAP in HeLa cell lysates (Figure 2.3B) and that it could actively promote the retrotransposition of RNA derived from the reporter plasmid ORF1*mneoI* (Figure 2.4B). Our data also suggested that free ORF1p is not expressed from p1K2_TAP and that the resulting ORF1p-ORF2p fusion peptide is not processed to liberate free ORF1p. Notably, we did not detect any evidence of an ORF1p-ORF2p fusion peptide from cells transfected with pAD2TE1 with either anti-ORF1p or anti-TAP antibodies (Figure 2.3B). Thus, these data are consistent with previous data suggesting that L1 translation does not produce an ORF1p-ORF2p fusion peptide (Leibold et al. 1990, McMillan and Singer 1993, Ergun et al. 2004, Alisch et al. 2006).

Methods

Cell culture

HeLa cells were grown in high-glucose DMEM (Gibco) supplemented with 10% FBS (Gibco), 100 U/mL penicillin-streptomycin (Invitrogen), and 0.29mg/mL L-glutamine (Gibco). Cell lines were maintained at 37°C with 7% CO2.

Plasmid Constructs

The following plasmids are based on the previously described pJM101/L1.3 construct (Sassaman et al. 1997). The amino acid and nucleotide numbers indicate the mutation position based on L1.3 accession number L19088 (Dombroski et al. 1993). The constructs were cloned into the pCEP4 expression vector (Invitrogen) and are equipped with the *mneoI* indicator cassette in the L1 3'UTR unless otherwise indicated. Sub-cloning was used to introduce the TAP epitope tag sequences onto the 3' end of ORF2. Because of this procedure, we deleted a portion of the L1 3'UTR (nucleotides 5818 to 5953). All plasmid DNA was prepared with a Midiprep Plasmid DNA Kit (QIAGEN).

pDK101 is identical to pDK101, but contains the $REKG_{235\text{-}238}$AAAA mutations in the ORF1p RRM domain (Moran et al. 1996, Kulpa and Moran 2005).

pDK105 is identical to pDK101, but contains the $RR_{261\text{-}262}$AA mutations in the ORF1p C-terminal domain (Moran et al. 1996, Wei et al. 2001, Kulpa and Moran 2005)

pDK106 is identical to pDK101, but contains the $RR_{261\text{-}262}$KK mutations in the ORF1p C-terminal domain (Moran et al. 1996, Kulpa and Moran 2005).

pDK108 is identical to pDK101, but contains the R_{262}K mutation in the ORF1p C-terminal domain (Moran et al. 1996, Kulpa and Moran 2005).

pDK109 is identical to pDK101, but contains the R_{261}K mutation in the ORF1p C-terminal domain(Moran et al. 1996, Kulpa and Moran 2005).

pDK116 is identical to pDK101, but contains the $YPAKLS_{282-287}AAAALA$ substitution in the ORF1p C-terminal domain (Moran et al. 1996, Kulpa and Moran 2005).

pDK135 is identical to pDK101, but contains the D702A mutation in the putative ORF2p RT active site (Moran et al. 1996, Wei et al. 2001, Kulpa and Moran 2005).

pDK136 is identical to pDK101, but contains the H230A mutation in the ORF2p EN domain (Moran et al. 1996, Wei et al. 2001, Kulpa and Moran 2005).

pLZC: is identical to pDK101, but contains four leucine to valine mutations ($L_{93,100,107,114}V$) in the ORF1p putative leucine zipper domain (Hulme 2007).

pLZC/AA is identical to pDK105, but also contains four leucine to valine mutations ($L_{93,100,107,114}V$; $RR_{261-262}AA$) (Hulme 2007).

pAD2TE1 is derived from pDK101 (L1.3) and contains both the T7 *gene 10* epitope tag on the carboxyl-terminus of ORF1p and a TAP tag on the carboxyl-terminus of ORF2p (Doucet et al. 2010).

pORF1*mneoI* consists of the L1 5' UTR, ORF1, and the spliced *mneoI* indicator cassette (Alisch et al. 2006).

p1K2_TAP is derived from GP1AAA2 (Alisch et al. 2006) and contains a TAP tag on the carboxyl-terminus of ORF2p. John B. Moldovan (University of Michigan Medical School) constructed the plasmid.

p1K2(RR-AA) is derived from p1K2_TAP, but contains the $RR_{261-262}AA$ mutations in the ORF1p C-terminal domain John B. Moldovan (University of Michigan Medical School) constructed the plasmid (Kulpa and Moran, 2005; Moran et al., 1996; Wei et al., 2001).

p1K2(LZC) is derived from p1K2_TAP, but contains four leucine to valine mutations ($L_{93,100,107,114}V$) in the ORF1p putative leucine zipper domain John B. Moldovan (University of Michigan Medical School) constructed the plasmid.

p1K2(LZC/AA) is derived from p1K2_TAP, but contains a putative leucine zipper domain as well as a C-terminal domain mutant ($L_{93,100,107,114}V$; $RR_{261-262}AA$) in ORF1p. John B. Moldovan (University of Michigan Medical School) constructed the plasmid.

p1K2(RF1) is derived from p1K2_TAP, except that the amino acids 1-255 have been deleted from ORF1p. John B. Moldovan (University of Michigan Medical School) constructed the plasmid.

p1K2(S119X) is derived from p1K2_TAP, but contains a stop codon instead of a serine 119 of ORF1p (Moran et al. 1996). John B. Moldovan (University of Michigan Medical School) constructed the plasmid.

pCEP/GFP, is a pCEP4-based plasmid that contains the humanized renilla green fluorescent protein (hrGFP) coding sequence from phrGFP-C (Stratagene) under the control of the pCEP4 CMV promoter (Alisch et al. 2006).

The L1 retrotransposition assay

The cultured cell retrotransposition assay was conducted as described previously (Moran et al. 1996, Wei et al. 2000). Briefly, ~$2x10^4$ HeLa cells/well were plated in 6 well dishes. The next day, each well was transfected with 1 µg of plasmid DNA using 3 µL of FuGENE 6 transfection reagent (Roche). Three days post-transfection, cells were grown in the presence of G418 (400 µg/mL) to select for retrotransposition events. The media was changed daily. After ~12 days of selection, the resultant cells were washed with 1X Phosphate-Buffered Saline (PBS), fixed, and stained with crystal violet to visualize colonies. In parallel, HeLa cells were plated in 6 well dishes and transfected with 0.5 µg of the same plasmids and pCEP4/GFP. Three days post-transfection cells were subjected to flow cytometry and the transfection efficiency was determined based on the number of GFP positive cells by FACS.

***Trans*-complementation assay**

The *trans*-complementation assay was performed as previously described (Wei et al. 2001). Briefly, HeLa cells were seeded in T-175 flasks (BD Falcon™)

at ~6×10^6 cells/flask and transfected the next day with 30 µg of plasmid DNA using 90 µL of FuGENE 6 transfection reagent (Roche). The next day, cells were co-transfected with equal amounts ORF1mneoI reporter plasmid and a driver L1 that lacked the mneoI indicator cassette. Parallel co-transfection experiments of both plasmids plus pCEP/GFP were used to determine the transfection efficiency of each sample (see above). Seventy-two hours post-transfection, the HeLa cells were subjected to G418 selection (400 µg/mL). After 12-14 d of selection, the cells were fixed as described above, and the resultant G418-resistant colonies were counted to determine the trans-complementation efficiency

Isolation of L1 RNPs and western blot analysis

Isolation of L1 RNPs has been described previously (Kulpa and Moran 2005, Kulpa and Moran 2006). Briefly, HeLa cells were seeded in T-175 flasks (BD Falcon™) at ~6×10^6 cells/flask and transfected the next day with 30 µg of plasmid DNA using 90 µL of FuGENE 6 transfection reagent. Two days post-transfection, hygromycin B (Gibco) (200 µg/mL) was added to the media to select for transfected cells. After one week of hygromycin selection, cells were harvested, lysed, and whole cell lysates were centrifuged at 1000 x g for 5 minutes. The cleared whole cell lysates were then centrifuged through an 8.5%/17% (w/v) sucrose cushion at 178,000 x g for 2 hours. The resultant pellet was resuspended with 100 µL dH2O + 1X Complete EDTA-free protease inhibitor cocktail (Roche). Bradford reagent (Bio-Rad) was used to determine protein concentration and the RNP sample was diluted to a final concentration of 1.5 mg/mL. Proteins were detected by western blot using the following primary antibodies: mouse anti-T7-Tag (Novagen), rabbit anti-TAP (Open Biosystems), mouse anti-α-tubulin (Sigma), rabbit anti-S6 (Cell Signaling Technology). Goat anti-mouse, anti-rabbit and anti-rat HRP-conjugated secondary antibodies were purchased from GE/Amersham. Western blots were developed using either the pico or femto ECL substrate (Pierce) according to manufacturer's protocols.

LEAP assay

The LEAP assay has been described previously(Kulpa and Moran 2006). Briefly, an aliquot (1 µL) of the RNP sample was added to 49 µL of LEAP assay master mix (50 mM Tris-HCL (pH=7.5), 50 mM KCl, 5 mM MgCl2, 10 mM DTT, 0.4 µM 3'RACE adapter primer, 20U RNasin (Promega), 0.2 mM dNTPs, and 0.05% (v/v) Tween 20) and was incubated at 37°C for 1 hour. LEAP cDNA products (1 µL) were amplified in a standard 50 µL PCR reaction containing 0.4 µM of the 3'RACE outer primer and 0.4 µM of one of the following forward primers: L1 3' end; Neo promoter sense; Neo8161S; LEAP-86; LEAP-46, using HotStart Pfu Turbo polymerase (Stratagene) according to the manufacturer's protocol. The resultant products were visualized on 2% agarose gels. LEAP PCR products were isolated, cloned into the pCR-Blunt vector (Invitrogen), and sequenced to confirm their identity.

RNA preparation and RT-PCR analysis

Total RNA was isolated using an RNeasy Kit (QIAGEN) coupled to an on-column DNase treatment (QIAGEN). The isolated RNAs were resuspended in Ultrapure distilled water (GIBCO) and quantified using a Nanodrop spectrophotometer (Thermo Scientific). For the LEAP assay controls, RNA was isolated from 50 µL RNP sample (1.5 mg/mL). RT-PCR was performed on 0.5 µg total RNA using M-MLV reverse transcriptase (Promega) followed by PCR with primers specific to the transfected L1 constructs or GAPDH. HotStart Pfu Turbo polymerase (Stratagene) then was used to amplify the resultant cDNA according to the manufacturer's instructions. The PCR protocol and the L1 and GAPDH oligonucleotide primers were previously described (Kulpa and Moran, 2006).

Northern Blots

HeLa cells were seeded in T-175 flasks (BD Falcon™) at ~6×10^6 cells/flask and transfected the next day with 30 µg of plasmid DNA using 90 µL of FuGENE 6 transfection reagent (Roche). Two days post-transfection, hygromycin B (Gibco) (200 µg/mL) was added to the media to select for transfected cells. After one week of hygromycin selection, cells were harvested, lysed, and RNA

was extracted with TRIzol® reagent (Ambion), and then poly(A)+ RNA was prepared from total RNA using an Oligotex mRNA kit (Qiagen). RNA samples were then subjected to glyoxal gel electrophoresis and northern blotting using the NorthernMax®-Gly Kit (Ambion) according to the manufactures protocol. Following electrophoresis, RNA was transferred to BrightStar® Nylon membranes (Invitrogen) and then cross-linked using UV light. For northern blot detection, membranes were prehybridized for ~4 hours at 68°C in NorthernMax® Prehybridization/Hybridization Buffer (Ambion), and then incubated with a strand specific RNA probe (final concentration of probe ~3×10^6 cpm ml-1) over night at 68°C.

Strand-specific RNA probes were generated using the MAXIscript® T3 system (Invitrogen). The 5UTR99 probe corresponds to bases 7-99 of the L1.3 5' UTR (Belancio et al. 2006) of the L1.3 sequence. RNA probe templates for T3 reactions were generated by PCR using pJM101/L1.3Δneo as a PCR template with the following primers:

5UTR99 (Forward): 5'-GGAGCCAAGATGGCCGAATAGGAACAGCT-3'

5UTR99 (Reverse): 5'-AATTAACCCTCAAAGGGACCTCAGATGGAAATGCAG-3'

The T3 promoter sequence (underlined) was added to the reverse primer of each primer pair. The pTRI-β-actin-125-Human Antisense Control Template (Applied Biosystems) was used in T3 reactions as a template to generate the β-actin RNA probe. The northern blot experiment was conducted one time.

Figure 2.1: Schematic diagram of engineered human L1 constructs.

The wild type human L1 construct, pDK101 expresses a human L1 (L1.3) (Sassaman et al. 1997) that contains a T7 *gene10* epitope tag (blue flag) on the carboxyl-terminus of ORF1p. The pDK101 construct also contains an *mneoI* retrotransposition indicator cassette located within the L1 3' UTR. A CMV promoter (black rectangle) augments L1 expression and an SV40 polyadenylation signal (pA) is located downstream of the native L1 polyadenylation signal. The *mneoI* cassette (red rectangle) is cloned into the L1 3' UTR antisense to the L1 and contains a neomycin phosphotransferase gene that is disrupted by an intron flanked by splice donor (SD) and splice acceptor sites (SA) in the L1 sense orientation. The neomycin phosphotransferase gene can only be expressed when the L1 transcript is spliced, reverse transcribed, and inserted into genomic DNA (Moran et al. 1996, Wei et al. 2000). The approximate locations of the ORF1 coiled coil (CC), RNA recognition motif (RRM), and carboxyl-terminal (CTD) domains are indicated. The approximate locations of the ORF2p endonuclease (EN) domain and the ORF2 reverse transcriptase (RT) domains are shown in the diagram. The approximate location of L1 mutations and the corresponding name of the mutant L1 plasmids are depicted on the diagram.

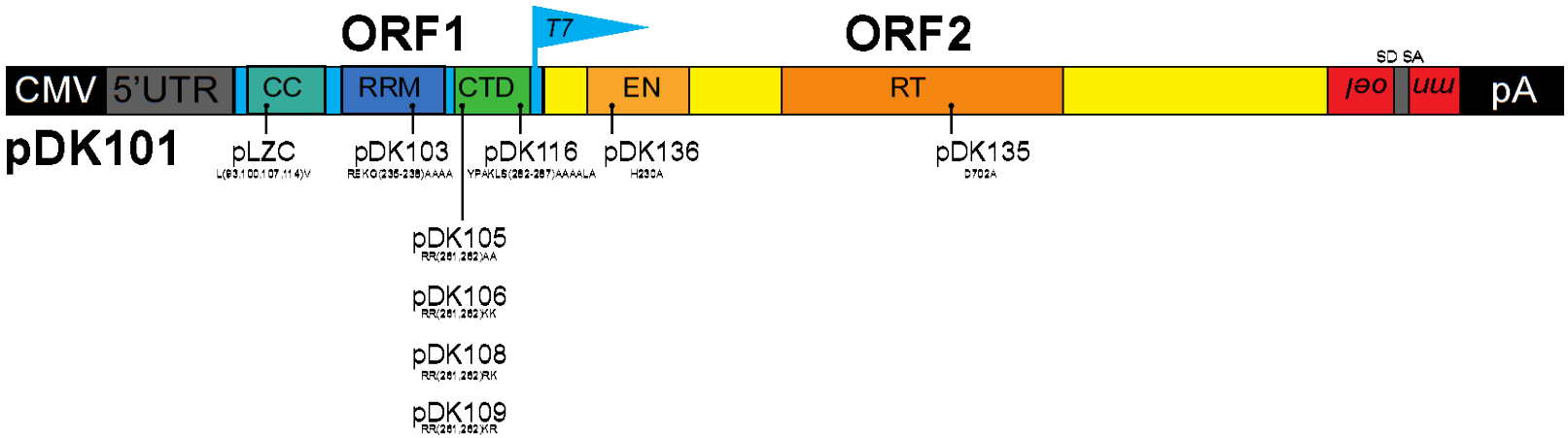


Figure 2.1: Schematic diagram of engineered human L1 constructs.

Figure 2.2: Mutations in ORF1p affect the formation of functional L1 RNPs.

A) *The effect of ORF1p mutations on L1 retrotransposition.* The X-axis indicates the name of the L1 construct. The Y-axis indicates retrotransposition efficiency normalized to the wild type L1 control, pDK101 (black bars). Error bars represent standard deviations based on a single experiment with three technical replicates. Beneath the graph is a single well of a representative six-well tissue culture plate, displaying G418-resistant colonies from retrotransposition assays. (B) *Mutations in ORF1p affect the localization of ORF1p to L1 RNPs.* Western blots of HeLa cells transfected with the indicated L1 plasmids (top of blots). Anti-T7 antibodies were used to detect the T7-tagged ORF1p expressed from L1 plasmids. Top panel: Western blots from transfected HeLa cell crude RNP fractions (RNPs). Bottom panel: Western blots from transfected HeLa cell lysates. UTF = untransfected HeLa cells. Ribosomal protein S6 was used as a loading control. The red box indicates portion of gel image previously published (Doucet et al. 2010). C) *LEAP assay results.* Top panel: Approximately 20 µL of LEAP reactions from the indicated crude L1 RNP preps (top of gels) was loaded onto agarose gels. pDK135 contains mutations in the ORF2p RT domain that render the element unable to retrotranspose, NoRNP = LEAP reaction did not contain RNPs; NTC = PCR water control. Middle panel: M-MLV RT-PCR with L1 LEAP PCR primers serves as a positive control for the presence of L1 RNA in RNP preps. Blue arrow indicates L1 cDNA band. Bottom panel: GAPDH RT-PCR served as a loading control. Molecular size ladders are to the left of gels. The red box indicates portion of gel image previously published (Doucet et al. 2010). I conducted all experiments shown in Figure 2.2.

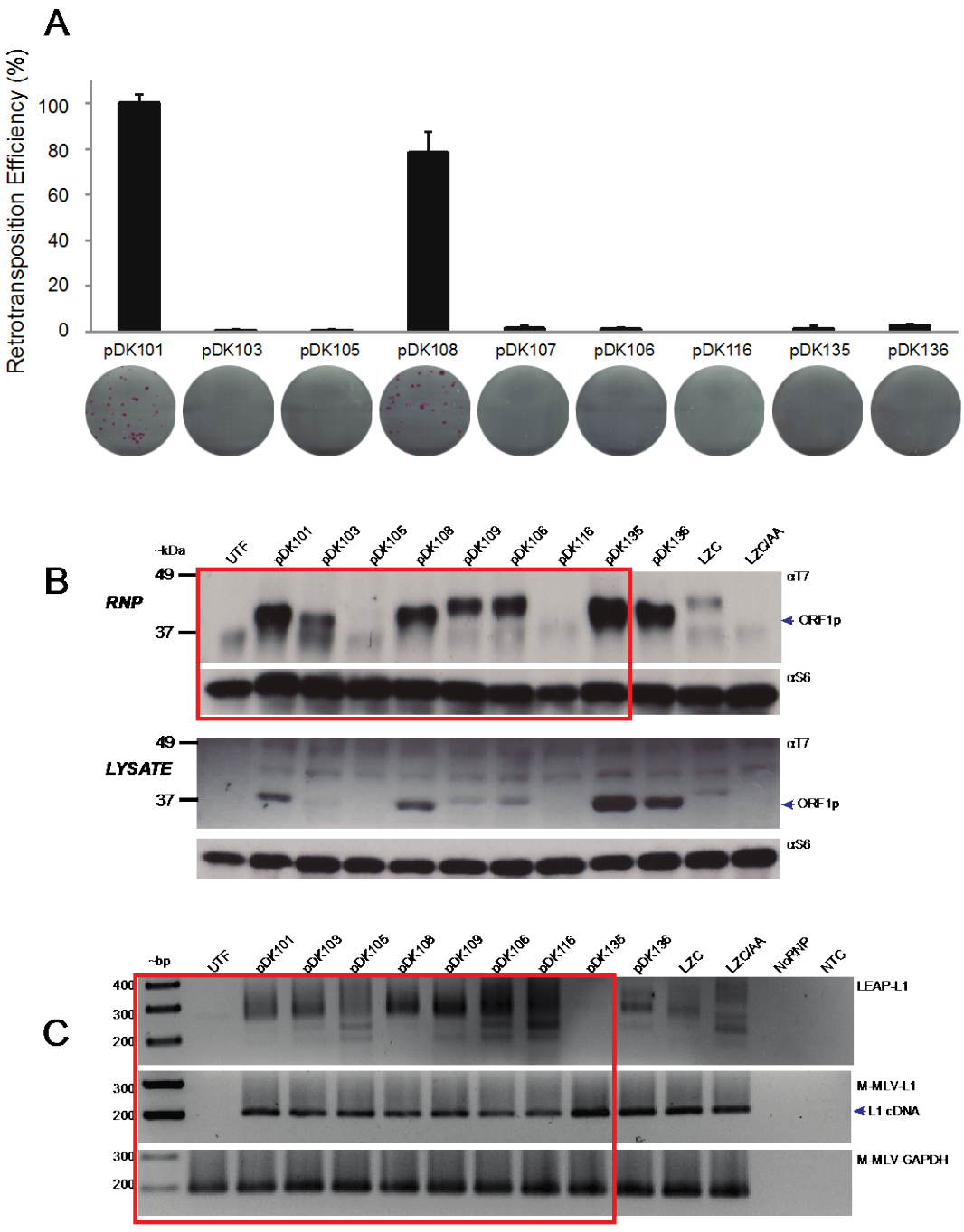


Figure 2.2: Mutations in ORF1p affect the formation of functional L1 RNPs.

Figure 2.3: Characterization of an L1 ORF1p-ORF2p fusion peptide.

A) *Schematic of L1 ORF1p-ORF2p fusion construct (p1K2_TAP).* p1K2_TAP was constructed by deleting the 63 bp spacer between ORF1 and ORF2 and mutating the ORF1 stop codon to a lysine (K) residue (green rectangle). The fusion construct is cloned into a pCEP4 mammalian expression vector and contains a TAP epitope tag (yellow flag) on the carboxyl-terminus of ORF2p to facilitate detection of the fusion peptide using anti-TAP antibodies. The expression of p1K2_TAP is augmented by the pCEP4 CMV promoter (black rectangle). The approximate locations of mutations with corresponding name of the mutant fusion constructs are mapped onto the p1K2_TAP diagram. B) *The ORF1p-ORF2p fusion peptide is expressed in HeLa cells.* HeLa cells were transfected with the indicated construct and RNPs were analyzed for the presence of the ORF1p-ORF2p fusion peptide using anti-TAP (top panel) or anti-ORF1p (bottom panel) antibodies. Green arrow indicates ORF1p-ORF2p fusion protein band; blue arrow indicates ORF2p-TAP protein band; red arrow indicates ORF1p-T7 protein band. Smaller bands in the p1K2_TAP lane of top panel may represent smaller TAP-tagged peptides (see main text) C) *Deletion of the amino-terminal amino acids (1-255) of ORF1p results in a proportional decrease in size of the 1K2 fusion peptide.* RNPs from transfected HeLa cells transfected with the indicated 1K2 fusion constructs (top of blots) were analyzed using anti-TAP antibodies. Red arrow indicates the approximate location of the p1K2(RF1) fusion peptide (~160 kDa). Tubulin was used as a loading control. D) *Northern blot results of 1K2 fusion RNA.* Top panel: Northern blots using a probe to the L1 5' UTR (5UTR99) was used to detect L1 RNA from poly-adenosine+ RNA extracted from HeLa cells transfected with the indicated constructs (above blots). Bottom panel: actin loading control northern blots. Blue arrow indicates L1 RNA. HeLa UTF = untransfected HeLa cells. E) *Mutations in ORF1p do not affect the expression of the ORF1p-ORF2p fusion peptide.* RNPs from transfected HeLa cells were analyzed with an anti-TAP antibody. Green arrow indicates ORF1p-ORF2p fusion protein band; blue arrow indicates ORF2p-TAP protein band. Tubulin was used as a loading control.

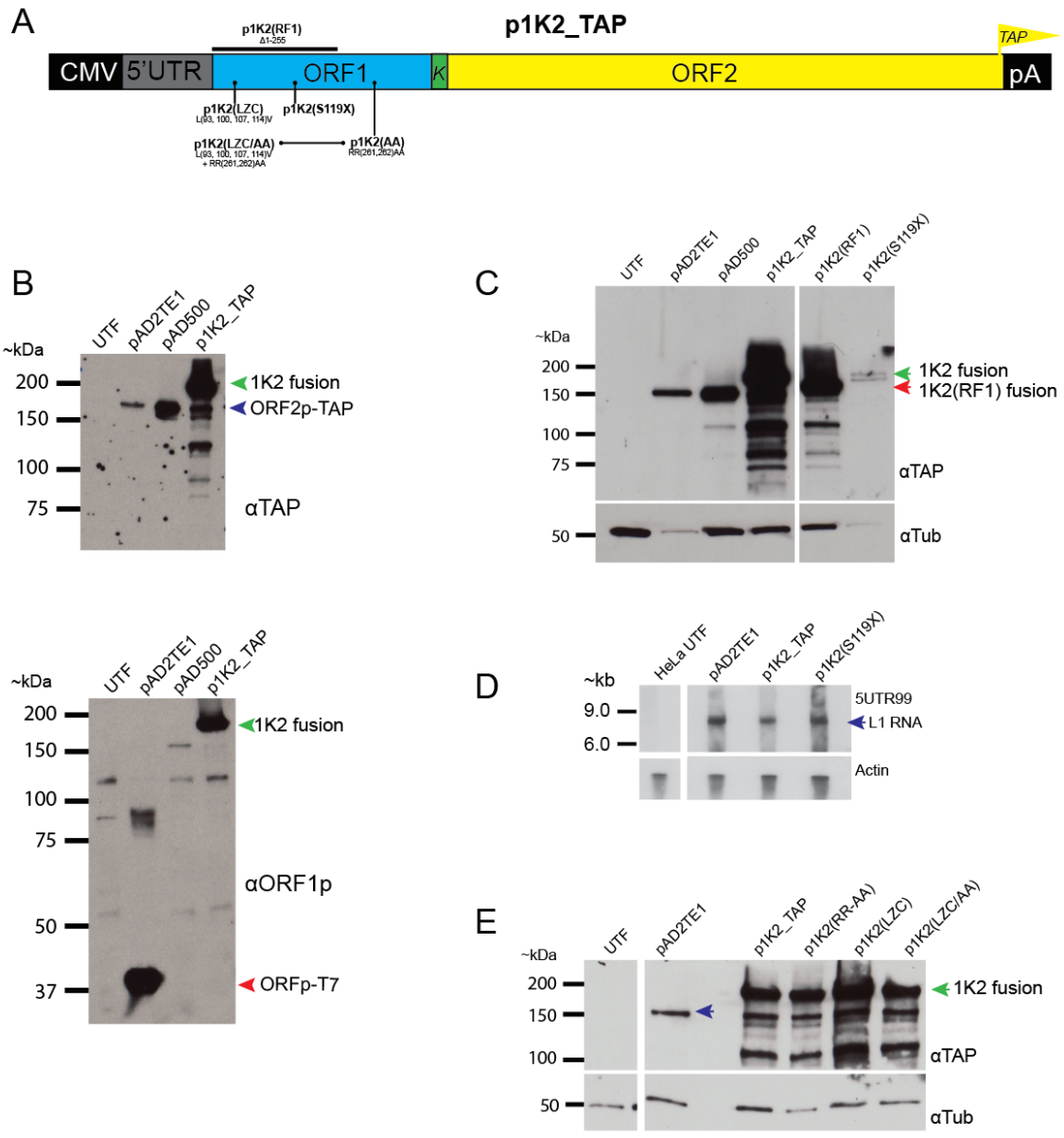


Figure 2.3: Characterization of an L1 ORF1p-ORF2p fusion peptide.

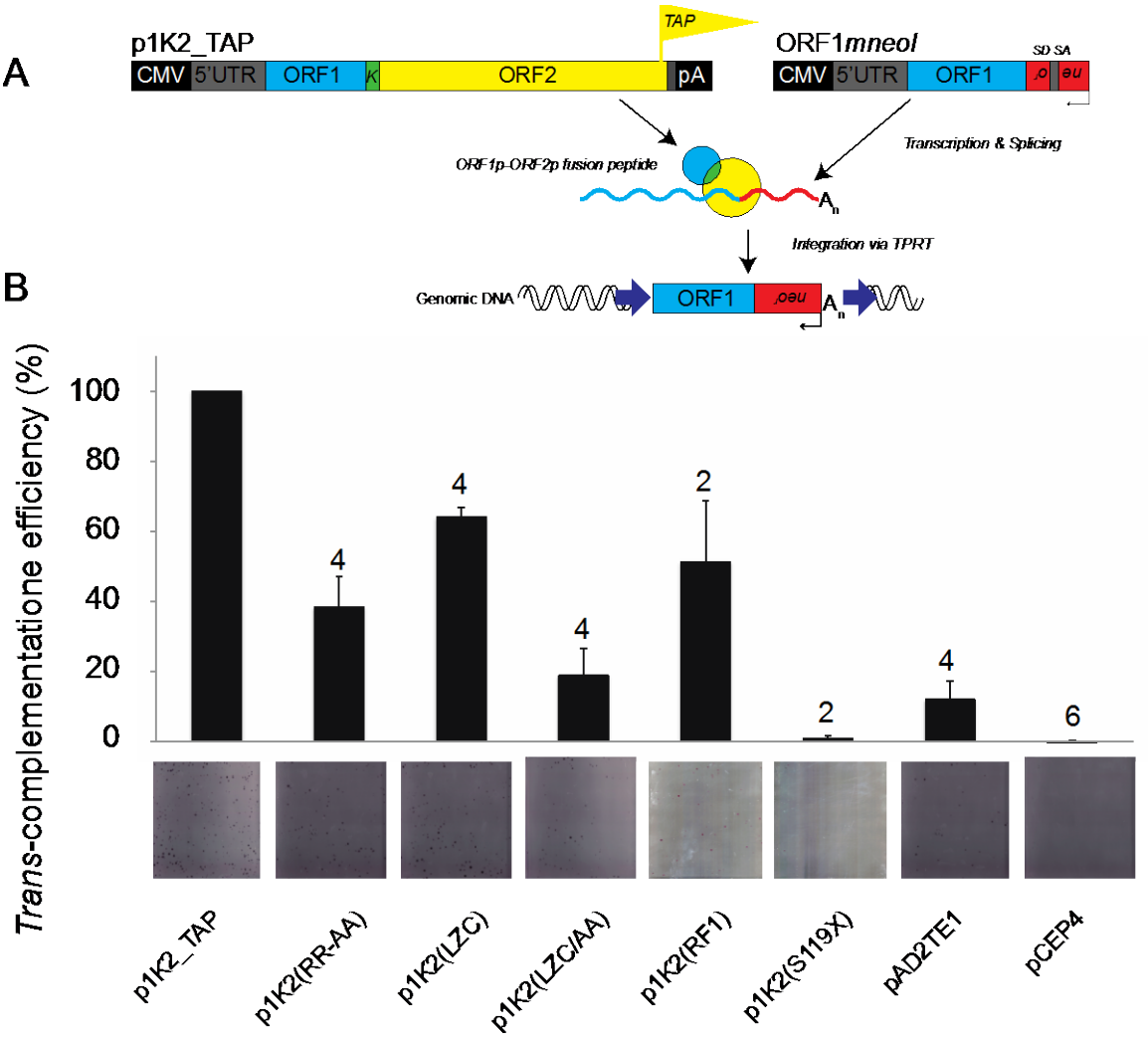


Figure 2.4: The effect of ORF1p mutations on 1K2 fusion *trans*-complementation efficiency.

(A) *Schematic of the trans-complementation assay.* HeLa cells are co-transfected with a driver L1 construct (*i.e.*, p1K2_TAP) and the ORF1*mneoI* reporter construct, which carries a neomycin retrotransposition indicator cassette (*mneoI*). L1 driver constructs are not marked with *mneoI* indicator cassette. Successful *trans*-complementation of ORF1*mneoI* will confer G418 resistance to transfected cells. (B) *Results of trans-complementation assay.* The X-axis indicates the L1 driver construct that was co-transfected with ORF1*mneoI*. The Y-axis on the graph indicates *trans*-complementation efficiency (black bars). *Trans*-complementation efficiency values have been normalized to p1K2_TAP wild type controls (set to 100%). The number above each bar indicates the number of independent experiments (n) performed with each L1 driver construct. Error bars represent standard deviations.

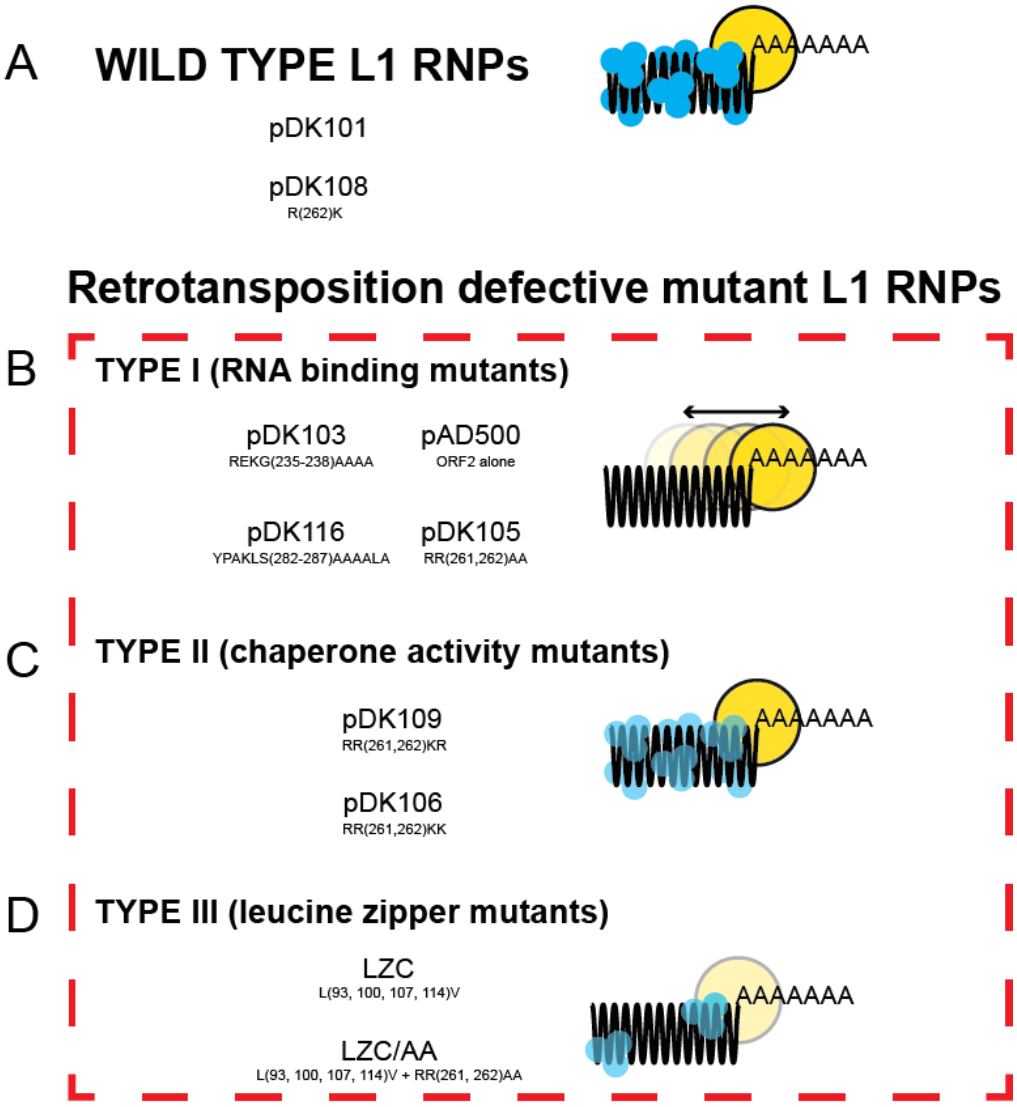


Figure 2.5: Classification of ORF1p mutant L1 RNPs.

A) *Wild type L1 RNPs (pDK101, pDK108):* Robust localization of L1 ORF1p (blue circles), ORF2p (yellow circle), and RNA (black wavy line) to RNPs; robust LEAP activity. B) *TYPE I (RNA binding mutants: pDK103, pDK105, pDK116, pAD500):* ORF1p fails to localize to L1 RNPs; RNPs contain wild type levels of ORF2p and exhibit wild type LEAP activity levels with an impairment in L1 RNA reverse transcription fidelity (*i.e.*, L1 reverse transcription can initiate from within internal L1 sequences). *C) TYPE II (chaperone activity mutants: pDK106, pDK109):* RNPs contain wild type levels of L1-encoded proteins and exhibit wild type LEAP activity. D) *TYPE III (leucine zipper mutants: pLZC, pLZC/AA):* Failure of ORF1p and ORF2p to localize to RNPs; decreased LEAP activity likely due to failure of ORF2p to localize to RNPs. Red-hatched box encloses L1 mutants that are not capable of retrotransposition.

References

Alisch, R. S., J. L. Garcia-Perez, A. R. Muotri, F. H. Gage and J. V. Moran, "Unconventional translation of mammalian LINE-1 retrotransposons." Genes Dev, 2006 **20**(2): 210-224.

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): 3846-3855.

Beck, C. R., P. Collier, C. Macfarlane, M. Malig, J. M. Kidd, E. E. Eichler, R. M. Badge and J. V. Moran, "LINE-1 retrotransposition activity in human genomes." Cell, 2010 **141**(7): 1159-1170.

Becker, K. G., G. D. Swergold, K. Ozato and R. E. Thayer, "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): 1697-1702.

Belancio, V. P., D. J. Hedges and P. Deininger, "LINE-1 RNA splicing and influences on mammalian gene expression." Nucleic Acids Research, 2006 **34**(5): 1512-1521.

Brouha, B., J. Schustak, R. M. Badge, S. Lutz-Prigge, A. H. Farley, J. V. Moran and H. H. Kazazian, Jr., "Hot L1s account for the bulk of retrotransposition in the human population." Proc Natl Acad Sci U S A, 2003 **100**(9): 5280-5285.

Callahan, K. E., A. B. Hickman, C. E. Jones, R. Ghirlando and A. V. Furano, "Polymerization and nucleic acid-binding properties of human L1 ORF1 protein." Nucleic Acids Res, 2012 **40**(2): 813-827.

Cordaux, R. and M. A. Batzer, "The impact of retrotransposons on human genome evolution." Nat Rev Genet, 2009 **10**(10): 691-703.

Dmitriev, S. E., D. E. Andreev, I. M. Terenin, I. A. Olovnikov, V. S. Prassolov, W. C. Merrick and I. N. Shatsky, "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): 4685-4697.

Dombroski, B. A., Q. Feng, S. L. Mathias, D. M. Sassaman, A. F. Scott, H. H. Kazazian, Jr. and J. D. Boeke, "An in vivo assay for the reverse transcriptase of human retrotransposon L1 in Saccharomyces cerevisiae." Mol Cell Biol, 1994 **14**(7): 4485-4492.

Dombroski, B. A., S. L. Mathias, E. Nanthakumar, A. F. Scott and H. H. Kazazian, Jr., "Isolation of an active human transposable element." Science, 1991 **254**(5039): 1805-1808.

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): 6513-6517.

Doucet, A. J., A. E. Hulme, E. Sahinovic, D. A. Kulpa, J. B. Moldovan, H. C. Kopera, J. N. Athanikar, M. Hasnaoui, A. Bucheton, J. V. Moran and N. Gilbert, "Characterization of LINE-1 ribonucleoprotein particles." PLoS Genet, 2010 **6**(10): e1001150.

Ergun, S., C. Buschmann, J. Heukeshoven, K. Dammann, F. Schnieders, H. Lauke, F. Chalajour, N. Kilic, W. H. Stratling and G. G. Schumann, "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): 27753-27763.

Esnault, C., J. Maestre and T. Heidmann, "Human LINE retrotransposons generate processed pseudogenes." Nat Genet, 2000 **24**(4): 363-367.

Feng, Q., J. V. Moran, H. H. Kazazian, Jr. and J. D. Boeke, "Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition." Cell, 1996 **87**(5): 905-916.

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): 46, 48-49, 52.

Goodier, J. L., P. K. Mandal, L. Zhang and H. H. Kazazian, Jr., "Discrete subcellular partitioning of human retrotransposon RNAs despite a common mechanism of genome insertion." Hum Mol Genet, 2010 **19**(9): 1712-1725.

Grimaldi, G., J. Skowronski and M. F. Singer, "Defining the beginning and end of KpnI family segments." EMBO J, 1984 **3**(8): 1753-1759.

Hancks, D. C. and H. H. Kazazian, Jr., "Active human retrotransposons: variation and disease." Curr Opin Genet Dev, 2012 **22**(3): 191-203.

Hohjoh, H. and M. F. Singer, "Cytoplasmic ribonucleoprotein complexes containing human LINE-1 protein and RNA." EMBO J, 1996 **15**(3): 630-639.

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): 6034-6043.

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): 19765-19768.

Hulme, A. E. (2007). The role of the ORF1 protein leucine zipper domain and cellular host factors in LINE-1 retrotransposition, University of Michigan.

Januszyk, K., P. W. Li, V. Villareal, D. Branciforte, H. Wu, Y. Xie, J. Feigon, J. A. Loo, S. L. Martin and R. T. Clubb, "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): 24893-24904.

Kazazian, H. H., Jr., C. Wong, H. Youssoufian, A. F. Scott, D. G. Phillips and S. E. Antonarakis, "Haemophilia A resulting from de novo insertion of L1 sequences

represents a novel mechanism for mutation in man." Nature, 1988 **332**(6160): 164-166.

Khazina, E., V. Truffault, R. Buttner, S. Schmidt, M. Coles and O. Weichenrieder, "Trimeric structure and flexibility of the L1ORF1 protein in human L1 retrotransposition." Nat Struct Mol Biol, 2011 **18**(9): 1006-1014.

Khazina, E. and O. Weichenrieder, "Non-LTR retrotransposons encode noncanonical RRM domains in their first open reading frame." Proceedings of the National Academy of Sciences, 2009 **106**(3): 731-736.

Kopera, H. C., J. B. Moldovan, T. A. Morrish, J. L. Garcia-Perez and J. V. Moran, "Similarities between long interspersed element-1 (LINE-1) reverse transcriptase and telomerase." Proc Natl Acad Sci U S A, 2011, 2011 **108**(51): 20345-20350.

Kubo, S., M. C. Seleme, H. S. Soifer, J. L. Perez, J. V. Moran, H. H. Kazazian, Jr. and N. Kasahara, "L1 retrotransposition in nondividing and primary human somatic cells." Proc Natl Acad Sci U S A, 2006 **103**(21): 8036-8041.

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): 3237-3248.

Kulpa, D. A. and J. V. Moran, "Cis-preferential LINE-1 reverse transcriptase activity in ribonucleoprotein particles." Nat Struct Mol Biol, 2006 **13**(7): 655-660.

Lander, E. S., L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh, R. Funke, D. Gage, K. Harris, A. Heaford, J. Howland, L. Kann, J. Lehoczky, R. LeVine, P. McEwan, K. McKernan, J. Meldrim, J. P. Mesirov, C. Miranda, W. Morris, J. Naylor, C. Raymond, M. Rosetti, R. Santos, A. Sheridan, C. Sougnez, N. Stange-Thomann, N. Stojanovic, A. Subramanian, D. Wyman, J. Rogers, J. Sulston, R. Ainscough, S. Beck, D. Bentley, J. Burton, C. Clee, N. Carter, A. Coulson, R. Deadman, P. Deloukas, A. Dunham, I. Dunham, R. Durbin, L. French, D. Grafham, S. Gregory, T. Hubbard, S. Humphray, A. Hunt, M. Jones, C. Lloyd, A. McMurray, L. Matthews, S. Mercer, S. Milne, J. C. Mullikin, A. Mungall, R. Plumb, M. Ross, R. Shownkeen, S. Sims, R. H. Waterston, R. K. Wilson, L. W. Hillier, J. D. McPherson, M. A. Marra, E. R. Mardis, L. A. Fulton, A. T. Chinwalla, K. H. Pepin, W. R. Gish, S. L. Chissoe, M. C. Wendl, K. D. Delehaunty, T. L. Miner, A. Delehaunty, J. B. Kramer, L. L. Cook, R. S. Fulton, D. L. Johnson, P. J. Minx, S. W. Clifton, T. Hawkins, E. Branscomb, P. Predki, P. Richardson, S. Wenning, T. Slezak, N. Doggett, J. F. Cheng, A. Olsen, S. Lucas, C. Elkin, E. Uberbacher, M. Frazier, R. A. Gibbs, D. M. Muzny, S. E. Scherer, J. B. Bouck, E. J. Sodergren, K. C. Worley, C. M. Rives, J. H. Gorrell, M. L. Metzker, S. L. Naylor, R. S. Kucherlapati, D. L. Nelson, G. M. Weinstock, Y. Sakaki, A. Fujiyama, M. Hattori, T. Yada, A. Toyoda, T. Itoh, C. Kawagoe, H. Watanabe, Y. Totoki, T. Taylor, J. Weissenbach, R. Heilig, W. Saurin, F. Artiguenave, P. Brottier, T. Bruls, E. Pelletier, C. Robert, P. Wincker, D. R. Smith, L. Doucette-Stamm, M. Rubenfield, K. Weinstock, H. M. Lee, J. Dubois, A. Rosenthal, M. Platzer, G. Nyakatura, S. Taudien, A. Rump, H. Yang, J. Yu, J. Wang, G. Huang, J. Gu, L. Hood, L.

Rowen, A. Madan, S. Qin, R. W. Davis, N. A. Federspiel, A. P. Abola, M. J. Proctor, R. M. Myers, J. Schmutz, M. Dickson, J. Grimwood, D. R. Cox, M. V. Olson, R. Kaul, C. Raymond, N. Shimizu, K. Kawasaki, S. Minoshima, G. A. Evans, M. Athanasiou, R. Schultz, B. A. Roe, F. Chen, H. Pan, J. Ramser, H. Lehrach, R. Reinhardt, W. R. McCombie, M. de la Bastide, N. Dedhia, H. Blocker, K. Hornischer, G. Nordsiek, R. Agarwala, L. Aravind, J. A. Bailey, A. Bateman, S. Batzoglou, E. Birney, P. Bork, D. G. Brown, C. B. Burge, L. Cerutti, H. C. Chen, D. Church, M. Clamp, R. R. Copley, T. Doerks, S. R. Eddy, E. E. Eichler, T. S. Furey, J. Galagan, J. G. Gilbert, C. Harmon, Y. Hayashizaki, D. Haussler, H. Hermjakob, K. Hokamp, W. Jang, L. S. Johnson, T. A. Jones, S. Kasif, A. Kaspryzk, S. Kennedy, W. J. Kent, P. Kitts, E. V. Koonin, I. Korf, D. Kulp, D. Lancet, T. M. Lowe, A. McLysaght, T. Mikkelsen, J. V. Moran, N. Mulder, V. J. Pollara, C. P. Ponting, G. Schuler, J. Schultz, G. Slater, A. F. Smit, E. Stupka, J. Szustakowski, D. Thierry-Mieg, J. Thierry-Mieg, L. Wagner, J. Wallis, R. Wheeler, A. Williams, Y. I. Wolf, K. H. Wolfe, S. P. Yang, R. F. Yeh, F. Collins, M. S. Guyer, J. Peterson, A. Felsenfeld, K. A. Wetterstrand, A. Patrinos, M. J. Morgan, P. de Jong, J. J. Catanese, K. Osoegawa, H. Shizuya, S. Choi and Y. J. Chen, "Initial sequencing and analysis of the human genome." Nature, 2001 **409**(6822): 860-921.

Leibold, D. M., G. D. Swergold, M. F. Singer, R. E. Thayer, B. A. Dombroski and T. G. Fanning, "Translation of LINE-1 DNA elements in vitro and in human cells." Proc Natl Acad Sci U S A, 1990 **87**(18): 6990-6994.

Luan, D. D., M. H. Korman, J. L. Jakubczak and T. H. Eickbush, "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): 595-605.

Martin, S. L., "Ribonucleoprotein particles with LINE-1 RNA in mouse embryonal carcinoma cells." Mol Cell Biol, 1991 **11**(9): 4804-4807.

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): 5383-5392.

Martin, S. L., D. Branciforte, D. Keller and D. L. Bain, "Trimeric structure for an essential protein in L1 retrotransposition." Proc Natl Acad Sci U S A, 2003 **100**(24): 13815-13820.

Martin, S. L., D. Bushman, F. Wang, P. W. Li, A. Walker, J. Cummiskey, D. Branciforte and M. C. Williams, "A single amino acid substitution in ORF1 dramatically decreases L1 retrotransposition and provides insight into nucleic acid chaperone activity." Nucleic Acids Res, 2008 **36**(18): 5845-5854.

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): 467-475.

Martin, S. L., M. Cruceanu, D. Branciforte, P. Wai-Lun Li, S. C. Kwok, R. S. Hodges and M. C. Williams, "LINE-1 retrotransposition requires the nucleic acid chaperone activity of the ORF1 protein." J Mol Biol, 2005 **348**(3): 549-561.

Mathias, S. L., A. F. Scott, H. H. Kazazian, Jr., J. D. Boeke and A. Gabriel, "Reverse transcriptase encoded by a human transposable element." Science, 1991 **254**(5039): 1808-1810.

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): 11533-11537.

Moran, J. V., S. E. Holmes, T. P. Naas, R. J. DeBerardinis, J. D. Boeke and H. H. Kazazian, Jr., "High frequency retrotransposition in cultured mammalian cells." Cell, 1996 **87**(5): 917-927.

Narita, N., H. Nishio, Y. Kitoh, Y. Ishikawa, R. Minami, H. Nakamura and M. Matsuo, "Insertion of a 5' truncated L1 element into the 3' end of exon 44 of the dystrophin gene resulted in skipping of the exon during splicing in a case of Duchenne muscular dystrophy." The Journal of Clinical Investigation, 1993 **91**(5): 1862-1867.

Richardson, S. R., A. J. Doucet, H. C. Kopera, J. B. Moldovan, J. L. Garcia-Perez and J. V. Moran, "The Influence of LINE-1 and SINE Retrotransposons on Mammalian Genomes." Microbiology Spectrum, 2015 **3**(2).

Sassaman, D. M., B. A. Dombroski, J. V. Moran, M. L. Kimberland, T. P. Naas, R. J. DeBerardinis, A. Gabriel, G. D. Swergold and H. H. Kazazian, Jr., "Many human L1 elements are capable of retrotransposition." Nat Genet, 1997 **16**(1): 37-43.

Scott, A. F., B. J. Schmeckpeper, M. Abdelrazik, C. T. Comey, B. O'Hara, J. P. Rossiter, T. Cooley, P. Heath, K. D. Smith and L. Margolet, "Origin of the human L1 elements: proposed progenitor genes deduced from a consensus DNA sequence." Genomics, 1987 **1**(2): 113-125.

Swergold, G. D., "Identification, characterization, and cell specificity of a human LINE-1 promoter." Mol Cell Biol, 1990 **10**(12): 6718-6729.

Taylor, M. S., J. Lacava, P. Mita, K. R. Molloy, C. R. Huang, D. Li, E. M. Adney, H. Jiang, K. H. Burns, B. T. Chait, M. P. Rout, J. D. Boeke and L. Dai, "Affinity Proteomics Reveals Human Host Factors Implicated in Discrete Stages of LINE-1 Retrotransposition." Cell, 2013 **155**(5): 1034-1048.

Wei, W., N. Gilbert, S. L. Ooi, J. F. Lawler, E. M. Ostertag, H. H. Kazazian, J. D. Boeke and J. V. Moran, "Human L1 retrotransposition: cis preference versus trans complementation." Mol Cell Biol, 2001 **21**(4): 1429-1439.

Wei, W., T. A. Morrish, R. S. Alisch and J. V. Moran, "A transient assay reveals that cultured human cells can accommodate multiple LINE-1 retrotransposition events." Anal Biochem, 2000 **284**(2): 435-438.

Xie, Y., L. Mates, Z. Ivics, Z. Izsvak, S. L. Martin and W. An, "Cell division promotes efficient retrotransposition in a stable L1 reporter cell line." Mob DNA, 2013 **4**(1): 10.

Chapter 3

The zinc-finger antiviral protein ZAP inhibits LINE and Alu retrotransposition

The research in this chapter was recently published as an article in *PLOS Genetics* (Moldovan and Moran 2015), and is presented here with only minor modifications to meet thesis formatting requirements. Use of the original article is in strict accordance with PLoS journal policies and the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/). I designed and carried out all experiments described in this chapter.

Abstract

Long INterspersed Element-1 (LINE-1 or L1) is the only active autonomous retrotransposon in the human genome. To investigate the interplay between the L1 retrotransposition machinery and the host cell, we used co-immunoprecipitation in conjunction with liquid chromatography and tandem mass spectrometry to identify cellular proteins that interact with the L1 first open reading frame-encoded protein, ORF1p. We identified 39 ORF1p-interacting candidate proteins including the zinc-finger antiviral protein (ZAP or ZC3HAV1). Here we show that the interaction between ZAP and ORF1p requires RNA and that ZAP overexpression in HeLa cells inhibits the retrotransposition of engineered human L1 and Alu elements, an engineered mouse L1, and an engineered zebrafish LINE-2 element. Consistently, siRNA-mediated depletion of endogenous ZAP in HeLa cells led to an ~2-fold increase in human L1 retrotransposition. Fluorescence microscopy in cultured human cells demonstrated that ZAP co-localizes with L1 RNA, ORF1p, and stress granule

associated proteins in cytoplasmic foci. Finally, molecular genetic and biochemical analyses indicate that ZAP reduces the accumulation of full-length L1 RNA and the L1-encoded proteins, yielding mechanistic insight about how ZAP may inhibit L1 retrotransposition. Together, these data suggest that ZAP inhibits the retrotransposition of LINE and Alu elements.

Introduction

Long INterspersed Element-1 (LINE-1, also known as L1) sequences comprise ~17% of human DNA and represent the only class of autonomously active retrotransposons in the genome (Lander et al. 2001). L1s mobilize (*i.e.*, retrotranspose) throughout the genome via an RNA intermediate by a copy-and-paste mechanism known as retrotransposition (reviewed in Beck et al. 2011). The overwhelming majority of human L1s are retrotransposition-deficient because they are 5' truncated, contain internal rearrangements (*i.e.*, inversion/deletion events), or harbor point mutations that compromise the functions of the L1-encoded proteins (ORF1p and ORF2p) (Grimaldi et al. 1984, Lander et al. 2001). Despite these facts, it is estimated that the average diploid human genome contains ~80-100 L1 elements that are capable of retrotransposition (Sassaman et al. 1997, Brouha et al. 2003, Beck et al. 2010). It is estimated that a new L1 insertion occurs in approximately 1 out of 200 live human births (reviewed in Cordaux and Batzer 2009). On occasion, L1 retrotransposition events can disrupt gene expression, leading to diseases such as hemophilia A (Kazazian et al. 1988), Duchenne muscular dystrophy (Holmes et al. 1994), and cancer (Miki et al. 1992, Shukla et al. 2013). Indeed, L1-mediated retrotransposition events are responsible for at least 96 disease-producing insertions in man (reviewed in Hancks and Kazazian 2012).

A full-length human L1 is ~6 kb in length and encodes a 5' UTR that harbors an internal RNA polymerase II promoter that directs transcription from at or near the first base of the element (Swergold 1990, Becker et al. 1993, Athanikar et al. 2004). The 5' UTR is followed by two open reading frames

(ORFs) that are separated by a short 63 bp inter-ORF spacer, and a 3' UTR that ends in a variable length poly adenosine (poly(A)) tract (Scott et al. 1987, Dombroski et al. 1991). The first L1 ORF encodes an ~40 kDa protein (ORF1p) that has nucleic acid binding (Martin 1991, Holmes et al. 1992, Hohjoh and Singer 1996, Hohjoh and Singer 1997, Khazina and Weichenrieder 2009) and nucleic acid chaperone activities (Martin and Bushman 2001, Khazina and Weichenrieder 2009). The second L1 ORF encodes a much larger ~150 kDa protein (ORF2p) (Ergun et al. 2004, Doucet et al. 2010, Goodier et al. 2010), which exhibits single-strand endonuclease (EN) (Feng et al. 1996) and reverse transcriptase (RT) (Mathias et al. 1991, Dombroski et al. 1994) activities. Experiments in cultured cells have revealed that activities associated with both ORF1p and ORF2p are required for efficient L1 retrotransposition (Feng et al. 1996, Moran et al. 1996).

During a cycle of L1 retrotransposition, a full-length L1 is transcribed and the resultant bicistronic L1 mRNA is exported to the cytoplasm where it undergoes translation. Notably, L1 RNA is translated in a cap-dependent manner by an unconventional termination-reinitiation mechanism that facilitates translation of both L1 ORFs (Leibold et al. 1990, McMillan and Singer 1993, Alisch et al. 2006, Dmitriev et al. 2007). Following translation, ORF1p and ORF2p preferentially bind to their respective encoding L1 mRNA template (a phenomenon known as *cis*-preference (Esnault et al. 2000, Wei et al. 2001)) to form an L1 ribonucleoprotein particle (RNP) (Martin 1991, Hohjoh and Singer 1996, Kulpa and Moran 2005, Kulpa and Moran 2006, Doucet et al. 2010). Components of the L1 RNP gain access to the nucleus by a process that does not strictly require cell division (Kubo et al. 2006), although L1 retrotransposition seems to be enhanced in dividing cells (Shi et al. 2007, Xie et al. 2013). Once the L1 RNP has entered the nucleus, the L1 RNA is reverse transcribed and inserted into genomic DNA by a process known as target-site primed reverse transcription (TPRT) (Luan et al. 1993, Feng et al. 1996, Cost et al. 2002). Briefly, the ORF2p endonuclease generates a single-strand endonucleolytic nick in genomic DNA at a thymidine rich consensus sequence (*e.g.*, 5'-TTTT/A, 5'-

TCTT/A, 5'-TTTA/A, *etc.*) (Feng et al. 1996, Cost and Boeke 1998, Morrish et al. 2002). The resulting 3' hydroxyl group then is used by the ORF2p reverse transcriptase as a primer to initiate (-) strand L1 cDNA synthesis from the L1 mRNA template (Feng et al. 1996, Cost and Boeke 1998). The completion of L1 integration requires elucidation, but likely involves host proteins involved in DNA repair and/or replication (Morrish et al. 2002, Gilbert et al. 2005, Suzuki et al. 2009, Taylor et al. 2013). Notably, the L1-encoded proteins also can work *in trans* to retrotranspose other cellular RNAs such as Short Interspersed Elements (SINEs) (*e.g.*, Alu (Dewannieux et al. 2003) and SINE-R/VNTR/Alu (SVA) elements (Ostertag et al. 2003, Hancks et al. 2011, Raiz et al. 2012)). L1 also can mobilize uracil-rich small nuclear RNAs (*e.g.*, U6 snRNA (Buzdin et al. 2002, Gilbert et al. 2005, Garcia-Perez et al. 2007), small nucleolar RNAs (*e.g.*, U3 snoRNA (Weber 2006)), and messenger RNAs, which results in the formation of processed pseudogenes (Esnault et al. 2000, Wei et al. 2001)).

Since L1 retrotransposition can be mutagenic, it stands to reason that the host cell employs multiple mechanisms to restrict L1 mobilization (reviewed in Levin and Moran 2011). For example, cytosine methylation of the L1 5' UTR suppresses L1 expression (Yoder et al. 1997, Bourc'his and Bestor 2004). In addition, piwi-interacting RNAs (piRNAs) suppress L1 expression in germ line cells (Aravin et al. 2007, reviewed in Levin and Moran 2011, and Siomi et al. 2011). Finally, emerging studies have demonstrated that several cellular proteins restrict L1 retrotransposition. These proteins include several APOBEC3 family members (reviewed in Schumann 2007, Richardson et al. 2014), TREX1 (Stetson et al. 2008), MOV10 (Arjan-Odedra et al. 2012, Goodier et al. 2012, Li et al. 2013), hnRNPL (Peddigari et al. 2013), SAMHD1 (Zhao et al. 2013), RNase L (Zhang et al. 2014), and the melatonin receptor 1 (MT1) (deHaro et al. 2014).

To gain a more complete understanding of the interplay between the L1 retrotransposition machinery and the host cell, we used liquid chromatography-tandem mass spectrometry (LC-MS/MS) to identify proteins that co-immunoprecipitate with L1 ORF1p in HeLa cells, reasoning that some of these proteins may affect L1 retrotransposition. We next analyzed the effects of

ORF1p-interacting proteins on L1 retrotransposition by overexpressing a subset of them in a cultured cell retrotransposition assay (Moran et al. 1996, Wei et al. 2000). Here, we report that the zinc-finger antiviral protein ZAP (Gao et al. 2002) interacts with L1 RNPs and inhibits L1 retrotransposition in cultured cells. ZAP also inhibits human Alu retrotransposition and the retrotransposition of mouse and zebrafish LINE elements. Molecular genetic and biochemical analyses suggest that ZAP inhibits retrotransposition by suppressing the accumulation of full-length L1 RNA and L1-encoded proteins in the cell.

Results

Identification of L1 ORF1p-interacting proteins

To identify proteins that interact with L1 ORF1p, we transfected HeLa cells with a human L1 construct, pJM101/L1.3FLAG, which expresses a version of ORF1p containing a FLAG epitope at its carboxyl-terminus (ORF1p-FLAG) (Figure 3.1A). The pJM101/L1.3FLAG construct exhibits robust retrotransposition activity in HeLa cells, albeit at a lower efficiency (~50%) than the untagged L1 construct, pJM101/L1.3 (Figure 3.8A).

Briefly, HeLa cells were transfected with pJM101/L1.3FLAG or pJM101/L1.3, a similar construct that lacks the FLAG epitope sequence (Figure 3.1A). Whole cell lysates from transfected cells then were incubated with anti-FLAG coated agarose beads to immunoprecipitate ORF1p-FLAG (see Methods). Immunoprecipitated fractions were analyzed by SDS-PAGE and proteins were visualized by silver staining (Figure 3.1B). The analysis of silver-stained gels revealed a prominent band of ~40 kDa (the theoretical molecular weight of ORF1p) in the pJM101/L1.3FLAG immunoprecipitation lane (Figure 3.1B; asterisk), which was not apparent in the pJM101/L1.3 lane. Western blot analysis with an antibody specific to L1.3 ORF1p (amino acids 31-49) confirmed the enrichment of ORF1p-FLAG in the pJM101/L1.3FLAG lane (Figures 3.1C and 3.8B; bottom panel). We also observed a complex pattern of bands between ~25 kDa and ~150 kDa that was present in the pJM101/L1.3FLAG lane that was not

evident in the pJM101/L1.3 lane (Figure 3.1B; black vertical bars). A similar pattern of protein bands was produced on silver-stained gels from pJM101/L1.3FLAG immunoprecipitation reactions using different wash and/or lysis conditions (Figures 3.8B and 3.8C, respectively).

To determine the identity of cellular proteins that associated with ORF1p-FLAG, the bands from the lanes corresponding to the pJM101/L1.3FLAG and pJM101/L1.3 immunoprecipitation experiments were excised from SDS-PAGE gels and submitted for LC-MS/MS (see Methods). An LC-MS/MS-identified protein was selected as an ORF1p-interacting candidate if it met the following criteria: 1) the protein was unique to the pJM101/L1.3FLAG immunoprecipitation, and 2) the protein was identified by two or more unique peptide sequences (peptide error rate ≤ 0.05; protein probability ≥ 0.95) (Table 3.1 and Methods). Thirty-nine ORF1p-interacting protein candidates were identified that met these criteria (Table 3.1).

To confirm the interactions between LC-MS/MS-identified proteins and ORF1p-FLAG, we evaluated 13 of the 39 ORF1p-FLAG interacting proteins for which there were commercially available antibodies and/or cDNA expression clones. Western blot analyses confirmed that these proteins associated with ORF1p-FLAG (Figure 3.1D). The 13 ORF1p-interacting proteins are involved in a variety of cellular processes including antiviral defense (ZAP (Gao et al. 2002) and MOV10 (Burdick et al. 2010)), nonsense-mediated decay (UPF1 (Leeds et al. 1991)), RNA splicing (hnRNPL (Hui et al. 2003) and DHX9 (Kernan et al. 1991, Kuroda et al. 1991)), and transcription (PURA (Bergemann et al. 1992), CDK9 (Grana et al. 1994), and ILF3 (Kao et al. 1994)). Notably, gene ontology (Ashburner et al. 2000) and global analyses of RNA binding proteins in human cell lines (Baltz et al. 2012, Castello et al. 2012) revealed that the 13 validated ORF1p-FLAG interacting proteins are RNA binding proteins (RBPs). Consistently, immunoprecipitation experiments of ORF1p-FLAG conducted in the presence of RNaseA revealed that the association between ORF1p and each of the 13 ORF1p-interacting proteins was sensitive to RNase A treatment (Figure 3.1D). Thus, the majority of ORF1p-interacting proteins associate with ORF1p by

binding to L1 RNA and/or other RNAs present within the L1 RNP (Mandal et al. 2013).

ORF1p-interacting proteins restrict L1 retrotransposition in HeLa cells

We next investigated whether overexpression of nine of the validated ORF1p-interacting proteins, as well as nine unvalidated ORF1p-interacting proteins, affects L1 retrotransposition (Moran et al. 1996, Wei et al. 2000). Briefly, HeLa cells were co-transfected with a cDNA plasmid expressing one of the ORF1p-FLAG interacting proteins and an engineered human L1 construct (pJJ101/L1.3; (Kopera et al. 2011)) marked with a blasticidin retrotransposition indicator cassette (*mblastI*) (Figures 3.2A and 3.2B; top panel). The *mblastI* cassette contains an antisense copy of the blasticidin deaminase gene, which is cloned into the L1 3' UTR. The blasticidin deaminase gene also is interrupted by an intron in the same transcriptional orientation as L1. This arrangement ensures that the blasticidin deaminase gene is expressed only when the L1 transcript is spliced, reverse transcribed, and inserted into genomic DNA. The resulting blasticidin-resistant foci then provide a visual, quantitative readout of retrotransposition activity (Moran et al. 1996, Morrish et al. 2002).

To monitor potentially toxic side effects of cDNA overexpression, HeLa cells also were co-transfected in a parallel assay with a cDNA expression vector and a control plasmid (pcDNA6/TR) that expresses the blasticidin deaminase gene (Figure 3.2B; bottom panel). Following blasticidin selection, the resulting foci provide a visual, quantitative readout of the effect of cDNA overexpression on colony formation (Figure 3.2B; bottom panel). This control is essential to determine if a cDNA affects L1 retrotransposition or cell viability and/or growth.

We co-transfected HeLa cells with each of the 18 ORF1p-FLAG interacting candidates and pJJ101/L1.3 (Figure 3.2C). An empty pCEP4 vector that was co-transfected with pJJ101/L1.3 served as a normalization control (Figures 3.2B and 3.2C). As a negative control, we demonstrated that a plasmid that expresses the humanized renilla green fluorescence protein (pCEP/GFP) did not affect pJJ101/L1.3 retrotransposition. As a positive control, we demonstrated

that a plasmid that expresses human APOBEC3A (pK_A3A) reduced pJJ101/L1.3 retrotransposition to ~18% of control levels (Figure 3.2C), which is in agreement with previous studies (Bogerd et al. 2006, Chen et al. 2006, Muckenfuss et al. 2006, Richardson et al. 2014). Four of the cDNA-expressing plasmids that we tested (ZAP-S (ZAP short isoform), hnRNPL, MOV10, and PURA) each reduced pJJ101/L1.3 retrotransposition to less than 50% of pCEP4 control levels. Notably, ZAP-S (~30% of control), hnRNPL (~30% of control), MOV10 (~13% of control), and PURA (~10% of control) inhibited retrotransposition to levels similar to that of pK_A3A (~18% of control) (Figure 3.2C). By comparison, the majority of the cDNA-expressing plasmids (14/18) did not significantly affect pJJ101/L1.3 retrotransposition levels (less than 50% inhibition when compared to pCEP4 control levels) (Figure 3.2C). Thus, the data suggest that ZAP-S, hnRNPL, MOV10, and PURA inhibit L1 retrotransposition in cultured cells.

ZAP inhibits L1 retrotransposition

The above data (Figure 3.2C) imply that ZAP, hnRNPL, MOV10, and PURA may function as host factors that restrict L1 retrotransposition. Notably, hnRNPL (Peddigari et al. 2013), MOV10 (Arjan-Odedra et al. 2012, Goodier et al. 2012), and PURA (Goodier et al. 2013) previously were shown to inhibit L1 retrotransposition. However, the effect of ZAP on L1 retrotransposition has not been studied; thus, we sought to determine how ZAP inhibits L1 retrotransposition.

ZAP is a poly (ADP-ribose) polymerase (PARP) family member (Kerns et al. 2008) initially characterized as an antiviral protein that inhibits murine leukemia virus (MLV) replication in cultured rat cells (Gao et al. 2002). Previous studies identified two human ZAP isoforms that resulted from alternative splicing (Kerns et al. 2008) (Figure 3.3A; top panel). The long ZAP isoform (ZAP-L) is 902 amino acids in length and contains an amino-terminus CCCH zinc-finger domain and an inactive carboxyl-terminal PARP-like domain (Kerns et al. 2008). The short ZAP isoform (ZAP-S) is 699 amino acids in length and lacks the carboxyl-

terminal PARP-like domain (Kerns et al. 2008). The HA-tagged human ZAP-L isoform restricted pJJ101/L1.3 retrotransposition to ~40% of control levels (Figure 3.3A; black bars) and the human ZAP-S isoform restricted pJJ101/L1.3 retrotransposition to ~30% of control levels (Figures 3.2C and 3.3A; black bars). Notably, overexpression of ZAP-L or ZAP-S did not dramatically affect the ability of HeLa cells to form blasticidin-resistant colonies in pcDNA6/TR control assays (Figure 3.3A, white bars). Western blot control experiments confirmed the overexpression of ectopic ZAP-L and ZAP-S compared to untransfected controls ~48 hours post-transfection (Figures 3.9A and 3.9B). Thus, ZAP inhibits L1 retrotransposition in cultured cells and the ZAP-L PARP-like domain is not required for L1 restriction.

Putative ZAP orthologs are present in several species (Kerns et al. 2008); thus, we tested whether a rat ZAP cDNA (rZAP) (Gao et al. 2002), that is orthologous to human ZAP-S (Gao et al. 2002, Kerns et al. 2008) could restrict pJJ101/L1.3 retrotransposition. Overexpression of rZAP efficiently reduced retrotransposition to ~40% of control levels (Figure 3.3A; black bars). Thus, the ability to restrict L1 retrotransposition is not limited to human ZAP.

The ZAP zinc-finger domain is necessary and sufficient to inhibit L1 retrotransposition

The ZAP zinc-finger domain binds to RNA and is required for antiviral activity (Guo et al. 2004, Chen et al. 2012). To analyze the role of the ZAP zinc-finger domain in L1 restriction, we tested the effects of a truncated ZAP-S mutant that expresses the ZAP zinc-finger domain (ZAP-S/1-311; containing amino acids 1-311) as well as a ZAP-S mutant that lacks the zinc-finger domain (ZAP-S/Δ72-372; lacking amino acids 72-372) in pJJ101/L1.3 retrotransposition assays (Figure 3.3A; above graph). ZAP-S/1-311 restricted retrotransposition to ~10% of control levels (Figure 3.3A; black bars), whereas ZAP-S/Δ72-372 had little effect on retrotransposition (~80% of control levels) (Figure 3.3A; black bars). The overexpression of the wild type or mutant ZAP-S/Δ72-372 expression constructs did not adversely affect the ability of HeLa cells to form blasticidin-resistant

colonies in pcDNA6/TR control assays (Figure 3.3A; white bars). Notably, transfection with ZAP-S/1-311 resulted in an ~50% decrease in the ability of HeLa cells to form blasticidin-resistant colonies; however, this effect has been accounted for through normalization (Figure 3.2B) and thus is independent of the ability of ZAP-S/1-311 to restrict L1 retrotransposition. Indeed, similar off-target effects have been reported for A3A cDNA expressing plasmids in HeLa cell-based L1 retrotransposition assays (Richardson et al. 2014). Western blot control experiments revealed that wild-type ZAP-S and the two mutant ZAP-S isoforms were expressed at similar levels ~48 hours post-transfection (Figures 3.9B and 3.9C). Thus, the ZAP zinc-finger domain is necessary and sufficient to inhibit L1 retrotransposition.

ZAP restricts the retrotransposition of various non-LTR retrotransposons

To determine if ZAP-S was able to restrict other non-long terminal repeat (non-LTR) retrotransposons, we tested whether ZAP-S expression affected human Alu retrotransposition. Unlike L1, Alu is a 7SL-derived non-autonomous retrotransposon that does not encode its own proteins (Ullu and Tschudi 1984). Instead, Alu elements must parasitize L1 ORF2p *in trans* to mediate their retrotransposition (Dewannieux et al. 2003). Briefly, HeLa cells were co-transfected with a full-length L1 element (pJM101/L1.3Δneo), an Alu retrotransposition reporter plasmid (pAlu*neo*^{Tet}), and a ZAP-S expression plasmid. Notably, ZAP-S potently reduced Alu retrotransposition to ~25% of control levels (Figure 3.3B). In contrast, the expression of the L1 restriction-deficient ZAP-S/Δ72-372 mutant did not negatively affect Alu retrotransposition (Figure 3.3B). Thus, ZAP-S is able to restrict the mobility of the two most prolific retrotransposons present in the human genome.

We next tested if human ZAP-S could restrict the retrotransposition of a natural mouse L1 (pG_F21) (Goodier et al. 2001), a zebrafish LINE-2 (pZfL2-2) (Sugano et al. 2006), or a synthetic mouse L1 (pCEPsmL1) (Han and Boeke 2004) that has been extensively mutagenized to alter 24% of the nucleic acid sequence without disrupting amino acid sequence. Human ZAP-S inhibited the

retrotransposition of human L1 (pJM101/L1.3; ~43% of control levels), natural mouse L1 (pG_F21; ~24% of control levels), zebrafish L2 (pZfL2-2; ~19% of control levels), and synthetic mouse L1 (pCEPsmL1; ~70% of control levels) (Figure 3.3C). The restriction-defective ZAP-S mutant, ZAP-S/Δ72-372, did not significantly affect the retrotransposition activity of these retrotransposons (Figure 3.3C). Notably, the milder inhibition of ZAP-S on pCEPsmL1 may be due to the elevated efficiency of pCEPsmL1 retrotransposition, the increased steady-state level of pCEPsmL1 mRNA and proteins, and/or the GC-rich nature of pCEPsmL1 (Han and Boeke 2004). Thus, ZAP-mediated restriction of retrotransposition is not specific to human non-LTR retrotransposons.

Depletion of endogenous ZAP enhances L1 retrotransposition

To test if endogenous ZAP restricts L1 retrotransposition, we used small interfering RNA (siRNA) to deplete endogenous ZAP from HeLa cells. Following siRNA treatment, cells were transfected with an L1 plasmid (pLRE3-*mEGFPI*) tagged with an *EGFP* indicator cassette (*mEGFPI*), which allows retrotransposition activity to be detected by EGFP fluorescence (Ostertag et al. 2000). As a negative control, HeLa cells were transfected with the L1 retrotransposition-defective plasmid pJM111-LRE3-*mEGFPI*, which carries two missense mutations that adversely affect ORF1p RNA binding (Moran et al. 1996, Martin et al. 2005, Khazina and Weichenrieder 2009). Treatment of HeLa cells with an siRNA pool against ZAP resulted in an ~80% and ~90% reduction of ZAP-L and ZAP-S protein levels, respectively, when compared to HeLa cells treated with a non-targeting control siRNA pool (Figure 3.3D; top left panel). ZAP siRNA treatment led to an approximately two-fold increase in pLRE3-*mEGFPI* retrotransposition activity when compared to assays conducted in the presence of a control siRNA (Figures 3.3D; bottom panel and 3.9D). We further demonstrated that siRNA-mediated depletion of endogenous MOV10 (Figure 3.3D; top right panel) from HeLa cells resulted in an approximately two-fold increase in pLRE3-*mEGFPI* retrotransposition (Figures 3.3D; bottom panel and 3.9D), which is in agreement with previous studies (Arjan-Odedra et al. 2012,

Goodier et al. 2012). These data suggest that endogenous ZAP may restrict L1 retrotransposition.

ZAP-S inhibits the accumulation of full-length LINE-1 mRNA

To investigate how ZAP restricts L1 retrotransposition, we analyzed the effect of ZAP-S expression on the accumulation of the L1 RNA. HeLa cells were co-transfected with pJM101/L1.3Δneo and either ZAP-S or ZAP-S/Δ72-372. Polyadenylated RNA from whole cell extracts then was analyzed by northern blot using RNA probes complementary to sequences within the L1.3 5' UTR (5UTR99) and ORF2 (ORF2_5804) (Figure 3.4A). Co-transfection with ZAP-S resulted in a reduction of full-length polyadenylated L1 RNA levels (~13% of pCEP4 control) compared to cells co-transfected with either the restriction-defective ZAP-S/Δ72-372 (~47% compared to pCEP4 control) or an empty pCEP4 control vector (Figure 4B; black arrow in blot; black bars in graph). Interestingly, ZAP-S expression did not have a pronounced effect on the accumulation of smaller L1 RNA species, which may have resulted from cryptic splicing and/or premature polyadenylation (Figure 3.4B; top panel: blue and yellow arrows, bottom panel: blue and yellow bars) (Perepelitsa-Belancio and Deininger 2003, Belancio et al. 2006, Belancio et al. 2008). Finally, control experiments revealed that ectopic ZAP-S expression did not affect endogenous actin RNA levels (Figure 3.4B). Thus, ZAP-S expression reduces the accumulation of full-length L1 mRNA in cultured cells.

ZAP-S inhibits the accumulation of ORF1p and ORF2p

We next examined the effect of ZAP-S expression on the accumulation of ORF1p and ORF2p. We co-transfected HeLa cells with either ZAP-S or ZAP-S/Δ72-372 and the L1 plasmid, pJBM2TE1, which expresses an L1.3 element marked with a T7 *gene10* epitope tag on the carboxyl-terminus of ORF1p and a TAP epitope-tag on the carboxyl-terminus of ORF2p (Figure 3.4C). Following co-transfection, HeLa cells were treated with puromycin to select for cells expressing pJBM2TE1. Both whole cell lysates (WCL) and RNP fractions were

collected 5 days post-transfection and subjected to western blot analyses to monitor ORF1p and ORF2p expression levels.

Expression of ZAP-S led to a decrease in the level of ORF1p and ORF2p in both WCL and RNP fractions, whereas the expression of the restriction-defective ZAP-SΔ/72-372 mutant or an empty pcDNA3 vector did not dramatically affect ORF1p or ORF2p expression levels (Figure 3.4D). The reduction in ORF1p and ORF2p was most evident in the RNP fraction, likely because both ORF1p and ORF2p are enriched in RNPs (Hohjoh and Singer 1996, Hohjoh and Singer 1997, Kulpa and Moran 2005, Kulpa and Moran 2006, Doucet et al. 2010). Control experiments revealed that ZAP-S expression did not affect the level of eIF3 protein (Figure 3.4D) and that ZAP-S and ZAP-S/Δ72-372 are expressed at similar levels in whole cell lysates (Figure 3.4D: top WCL panel). By comparison, ZAP-SΔ/72-372 is present at much lower levels in the RNP fraction compared to wild-type ZAP-S (Figure 3.4D; bottom RNP panel), suggesting that the zinc-finger domain is responsible for ZAP-S localization to the RNP fraction.

To determine if ZAP-S affects the expression of non-L1 proteins, we examined the effect of ZAP-S on EGFP expression. We co-transfected ZAP-S with an L1 plasmid (pLRE3-EF1-*mEGFP*ΔIntron) (Wissing et al. 2011) that expresses the L1 element, LRE3 and an intact copy of the EGFP gene (Figure 3.10A). In this case, LRE3 and EGFP are under the control of convergent promoters, which allows the simultaneous expression of LRE3 and EGFP from pLRE3-EF1-*mEGFP*ΔIntron. Thus, EGFP expression is not dependent on retrotransposition. Forty-eight hours post-transfection, flow cytometry was used to isolate EGFP-positive cells (*i.e.,* cells expressing pLRE3-EF1-*mEGFP*ΔIntron) (Figure 3.10C). Western blotting demonstrated a marked reduction in ORF1p when compared to EGFP levels in cells that were co-transfected with ZAP-S (Figure 3.10B). By comparison, ORF1p and EGFP were present at comparable levels in cells that were co-transfected with either an empty pCEP4 vector or the restriction-deficient ZAP-SΔ/72-372 mutant (Figure 3.10B). Control experiments revealed that ZAP-S did not affect endogenous tubulin protein levels (Figure

3.10B). Thus, ZAP-S expression appears to preferentially restrict the expression of L1 ORF1p.

ZAP co-localizes with ORF1p and L1 RNA in the cytoplasm

ORF1p, ORF2p, and L1 RNA form RNP complexes that appear as discrete cytoplasmic foci when visualized by fluorescence microscopy (Goodier et al. 2007, Doucet et al. 2010, Goodier et al. 2010). Notably, previous studies have shown that ZAP predominantly is localized in the cytoplasm (Liu et al. 2004) and that ZAP antiviral activity also is localized to the cytoplasm (Gao et al. 2002). To determine if ZAP co-localizes with ORF1p, we co-transfected HeLa cells with pJM101/L1.3Δneo and a plasmid that expresses a carboxyl-terminus turbo-GFP tagged ZAP-S protein (ZAP-S-tGFP). Control experiments showed that ZAP-S-tGFP restricted pJJ101/L1.3 retrotransposition to ~55% of control levels (Figure 3.11A). Confocal fluorescence microscopy revealed that ORF1p and ZAP-S-tGFP co-localized in discrete cytoplasmic foci in ~68% of cells that co-expressed both ORF1p and ZAP-S-tGFP (Figure 3.5A). To test if transfected ORF1p co-localizes with endogenous ZAP, we transfected HeLa cells with pAD2TE1, which expresses a human L1 (L1.3) containing a T7 *gene10* epitope-tag on the carboxyl-terminus of ORF1p (Doucet et al. 2010). Confocal microscopy revealed that endogenous ZAP co-localized with ORF1p-T7 in cytoplasmic foci in ~91% of cells that contained ORF1p-T7 foci (Figure 3.5B). Next, to test if endogenous ORF1p co-localizes with transfected ZAP-S, we transfected PA-1 cells (a human embryonic carcinoma-derived cell line that expresses endogenous ORF1p (Zeuthen et al. 1980, Garcia-Perez et al. 2010)) with ZAP-S-tGFP. Confocal microscopy demonstrated that endogenous ORF1p co-localized with ZAP-S-tGFP in ~89% of PA-1 cells that expressed ZAP-S-tGFP foci (Figure 3.5C). Thus, ORF1p and ZAP generally localize to the same region of the cytoplasm.

To test if the ZAP-S zinc-finger domain is critical for the co-localization of ZAP-S with ORF1p, we co-transfected HeLa cells with pJM101/L1.3Δneo and a tGFP-tagged ZAP-S mutant that expresses the ZAP-S zinc-finger domain (ZAP-S/Δ310-645-tGFP; lacking amino acids 310-645), or a ZAP-S mutant that lacks

the ZAP-S zinc-finger domain (ZAP-S/Δ72-372-tGFP; lacking amino acids 72-372) (Figure 3.11A). In control experiments, ZAP-S/Δ310-645-tGFP restricted pJJ101/L1.3 retrotransposition to ~32% of control levels whereas ZAP-S/Δ72-372-tGFP did not have a significant effect (~93% of control) on retrotransposition activity (Figure 3.11A). Confocal microscopy revealed that ORF1p and ZAP-S/Δ310-645-tGFP co-localized in cytoplasmic foci in ~74% of cells that co-expressed both ORF1p and ZAP-S/Δ310-645-tGFP (Figure 3.5D). In cells transfected with pJM101/L1.3Δneo and ZAP-S/Δ72-372-tGFP, ORF1p and ZAP-S/Δ72-372-tGFP co-localized in only ~14% of cells that co-expressed both ORF1p and ZAP-S/Δ72-372-tGFP (Figure 3.5C). Thus, the ZAP-S zinc-finger domain is necessary and sufficient for the co-localization of ZAP-S and ORF1p in cytoplasmic foci.

To determine if ZAP co-localizes with L1 RNA, we co-transfected HeLa cells with pJM101/L1.3 and either ZAP-S-tGFP, ZAP-S/Δ310-645-tGFP, or ZAP-S/Δ72-372-tGFP. To visualize L1 RNA, transfected cells were probed with fluorescently labeled oligonucleotide probes complementary to sequences within the L1 5' UTR. As a control, cells were co-transfected with pJM101/L1.3 and an empty pCEP4 vector. In pCEP4 control experiments, fluorescence microscopy revealed that ORF1p co-localized with L1 RNA in cytoplasmic foci in ~88% of cells that contained ORF1p cytoplasmic Foci (Figure 3.6A and Figure 3.12A). In HeLa cells co-transfected with pJM101/L1.3 and ZAP-S-tGFP, L1 RNA co-localized with ORF1p and ZAP-S-tGFP in cytoplasmic foci in ~23% of foci-containing cells (Figure 3.6B and Figure 3.12A). Thus, ZAP and L1 RNA co-localize in cytoplasmic foci.

Fluorescence microscopy further revealed that in cells co-transfected with pJM101/L1.3 and ZAP-S/Δ310-645-tGFP that L1 RNA was detected in ORF1p and ZAP-S/Δ310-645-tGFP foci in only ~18% of foci-containing cells (Figure 3.6C and Figure 3.12A). In contrast, in cells co-transfected with pJM101/L1.3 and ZAP-S/Δ72-372-tGFP, L1 RNA co-localized with ORF1p in ~77% of cells that expressed ZAP/Δ72-372-tGFP and contained ORF1p cytoplasmic Foci (Figure

3.6D and Figure 3.12A). Thus, the data suggest that ZAP prevents the accumulation of L1 RNA in cytoplasmic foci.

We next determined the effect of ZAP-S on ORF1p expression using confocal microscopy. HeLa cells were co-transfected with pJM101/L1.3Δneo and either ZAP-S-tGFP, ZAP-S/Δ310-645-tGFP, or ZAP-S/Δ72-372-tGFP. As a control, cells were co-transfected with pJM101/L1.3Δneo and an empty pCEP4 vector. In pCEP4 control experiments, confocal microscopy revealed that ~10.8% of cells expressed ORF1p after ~48 hours (Figure 3.12B). In contrast, only ~2.8% of cells that were co-transfected with ZAP-S-tGFP expressed ORF1p and ~2.8% of cells that were co-transfected with ZAP-S/Δ310-645-tGFP expressed ORF1p (Figure 3.12B). Approximately 8.0% of cells that were co-transfected with ZAP-S/Δ72-372-tGFP expressed ORF1p (Figure 3.12B). Thus, the data suggest that the ZAP-S zinc-finger domain is necessary and sufficient to inhibit the accumulation of ORF1p in HeLa cells.

L1 cytoplasmic foci also co-localize with an array of RNA binding proteins, including markers of cytoplasmic stress granules (SGs) (Goodier et al. 2007, Doucet et al. 2010, Goodier et al. 2013). Notably, ZAP also localizes to cytoplasmic SGs (Leung et al. 2011). To determine whether ZAP-S/ORF1p foci co-localize with cytoplasmic SGs we transfected HeLa cells with pJM101/L1.3Δneo and ZAP-S-tGFP. Confocal microscopy revealed that ZAP-S-tGFP/ORF1p co-localized with the endogenous SG associated protein eIF3 (Figure 3.11B). Additionally, endogenous ZAP also co-localized with the SG marker, G3BP (Figure 3.11D). In contrast, ZAP-S-tGFP/ORF1p foci did not co-localize with endogenous tubulin (Figure 3.11C), and endogenous ZAP did not co-localize with the processing body associated protein, DCP1α (Figure 3.11E). Thus, L1 ORF1p, ZAP, and SG associated proteins partition to the same cytoplasmic compartment.

Discussion

In this study, we identified 39 cellular proteins that interact with L1 ORF1p and validated 13 of these interactions in biochemical assays. Our data showed that the 13 validated ORF1p-interacting proteins associate with ORF1p via an RNA bridge (Figure 3.1D). Notably, 33 out of 39 of the ORF1p-interacting proteins also were detected in recent studies (Table S1; (Goodier et al. 2013, Peddigari et al. 2013, Taylor et al. 2013)). Importantly, we discovered that ZAP restricts human L1 and Alu retrotransposition. We also showed that hnRNPL, MOV10, and PURA inhibit L1 retrotransposition, which is in agreement with previous studies (Arjan-Odedra et al. 2012, Goodier et al. 2012, Goodier et al. 2013, Li et al. 2013, Peddigari et al. 2013). Thus, our data both confirm and extend those previous analyses and will help guide future studies that endeavor to determine how L1 retrotransposition impacts the human genome.

ZAP inhibits the mobility of both human and non-human non-LTR retrotransposons. The overexpression of the human and rat orthologs of ZAP restricts human L1 retrotransposition (Figure 3.3A). Human ZAP-S overexpression restricts the retrotransposition of an engineered human SINE (Alu) (Figure 3.3B), an engineered mouse L1 (G_F21), and an engineered zebrafish LINE-2 element (ZfL2-2) (Figure 3.3C). Although our studies primarily involved the overexpression of ZAP, we also demonstrated that the depletion of endogenous ZAP in HeLa cells led to an ~2-fold increase in L1 retrotransposition (Figure 3.3D). This observed increase in L1 retrotransposition activity is similar to increases in L1 activity that were observed upon depletion of MOV10 and hnRNPL proteins in other studies (Arjan-Odedra et al. 2012, Goodier et al. 2012, Peddigari et al. 2013). Thus, in principle, physiological levels of ZAP may be sufficient to influence retrotransposition in certain cell types.

The ZAP CCCH zinc-finger domain is required to both bind and mediate the degradation of viral RNA (Guo et al. 2004, Guo et al. 2007). Our data indicate that ZAP binding to L1 RNA is critical for L1 restriction. We demonstrated that ORF1p-FLAG and ZAP interact via an RNA bridge (Figure

3.1D). Moreover, we showed that overexpression of the ZAP zinc-finger domain more potently inhibits L1 retrotransposition than overexpression of wild type ZAP-L or ZAP-S (Figure 3.3A), and that the ZAP-S zinc-finger domain is required to inhibit L1 retrotransposition (Figure 3.3A and Figure 3.11A). In addition to our genetic and biochemical data, fluorescence microscopy revealed that: 1) co-transfected ZAP-S, L1 RNA, and ORF1p co-localize in the cytoplasm of HeLa cells; 2) the ZAP-S zinc-finger domain is necessary and sufficient for the co-localization of ZAP-S, L1 RNA, and ORF1p; 3) endogenous ZAP co-localizes with transfected ORF1p in HeLa cells; and 4) endogenous ORF1p interacts with transfected ZAP-S in human PA-1 embryonic carcinoma cells (Figures 3.5A-3.5E and Figures 3.6A-3.6D). Thus, the data suggest that ZAP interacts with L1 RNA in order to mediate L1 restriction.

Notably, the zebrafish ZfL2-2 retrotransposon lacks a homolog to ORF1 and only encodes a single ORF that contains an apurinic/apyrimidinic endonuclease-like (EN) and a reverse transcriptase (RT) domain (Sugano et al. 2006). The finding that ZAP-S efficiently restricts ZfL2-2 retrotransposition further indicates that ZAP-S likely restricts retrotransposition by interacting with LINE RNA. Although a ZAP consensus RNA target sequence/motif has not yet been identified, evidence suggests that ZAP recognizes long RNA stretches (>500 nucleotides) and/or specific RNA tertiary structure (Guo et al. 2004, Chen et al. 2012). The ability of ZAP to inhibit non-human LINE elements suggests that ZAP may not recognize a particular LINE linear consensus RNA sequence, but instead may recognize an unidentified structural feature common to certain LINE RNAs (Guo et al. 2004, Chen et al. 2012).

Evidence suggests that ZAP prevents the accumulation of viral RNAs in the cytoplasm (Gao et al. 2002). ZAP-S overexpression significantly reduced the amount of polyadenylated, full-length L1 RNA (Figure 3.4B), which would be expected to inhibit retrotransposition by limiting the supply of L1 mRNA available for translation and as a template for TPRT. Notably, while ZAP-S selectively inhibited the accumulation of full-length L1 transcripts, it did not dramatically affect the accumulation of shorter, spliced and/or polyadenylated L1 RNAs

(Figure 3.4B) (Perepelitsa-Belancio and Deininger 2003, Belancio et al. 2006, Belancio et al. 2008, Belancio et al. 2010). Thus, ZAP does not appear to affect L1 transcription *per se*, but instead likely affects the post-transcriptional processing of full-length L1 mRNA. In addition to biochemical data, fluorescence microscopy revealed that L1 RNA was depleted from L1 ORF1p cytoplasmic foci in the presence of ZAP (Figure 3.11A). The depletion of RNA from L1 cytoplasmic foci was dependent on the ZAP-S zinc-finger domain. Based on these data it is likely that ZAP prevents the accumulation of cytoplasmic L1 mRNA.

Previous studies have shown that ZAP also suppresses the expression of viral proteins (Bick et al. 2003, Zhu et al. 2011, Zhu et al. 2012). Western blot experiments demonstrated that the overexpression of ZAP-S inhibited the accumulation of L1 ORF1p and L1 ORF2p in whole cell lysates and RNPs derived from transfected HeLa cells (Figure 3.4D and Figure 3.10B). In agreement with western blot experiments, confocal microscopy experiments showed that tGFP-tagged ZAP-S inhibited the expression of ORF1p in transfected HeLa cells and that the ZAP-S zinc-finger domain is critical for the inhibition of ORF1p expression (Figure 3.12B). ZAP-S expression also inhibited Alu retrotransposition (Figure 3.3B), which depends on ORF2p to be supplied *in trans* by L1 (Dewannieux et al. 2003). In contrast to these data, ZAP-S overexpression did not significantly affect the expression and/or accumulation of EGFP or other endogenous proteins (*e.g.*, eiF3 and tubulin) (Figure 3.4D and Figure 3.10B). Thus, ZAP may preferentially limit the accumulation of ORF1p and ORF2p by interacting with L1 mRNA.

In sum, our data suggest that ZAP restricts L1 retrotransposition by preventing the accumulation of cytoplasmic L1 RNA. Notably, a recent study suggests that ZAP may interfere with translation of viral RNA, and that translation inhibition may precede viral RNA destruction (Zhu et al. 2012). Although a reduction in L1 RNA could explain the observed decrease in L1 protein expression, it also is conceivable that the interaction between ZAP and L1 RNA could interfere with L1 translation (Figure 3.7).

It remains unclear how ZAP might destabilize full-length L1 RNA to restrict retrotransposition. Evidence suggests that ZAP recruits exosome components (Guo et al. 2007) along with other proteins involved in RNA degradation (Chen et al. 2008, Zhu et al. 2011) to destroy viral RNA. Interestingly, immunofluorescence microscopy experiments revealed that ORF1p and ZAP co-localize with components of cytoplasmic SGs (Figures 3.10B and 3.10D), which contain numerous RNA binding proteins involved in cytosolic RNA metabolism (reviewed in Buchan and Parker 2009). Indeed, ZAP previously has been shown to localize to SGs (Leung et al. 2011) and SGs have been suggested to play a role in regulating L1 retrotransposition (Goodier et al. 2007) and viral pathogenesis (Reineke and Lloyd 2013). The co-localization of ORF1p and ZAP with SGs also suggests that ZAP may possibly inhibit L1 translation, as SG assembly is stimulated by translational arrest (reviewed in Buchan and Parker 2009). Thus, we propose that ZAP interacts directly with L1 RNA in the cytoplasm, which likely results in the recruitment in SG components and/or other cellular factors involved in RNA metabolism to destabilize L1 RNA and/or block translation (Figure 3.7).

ZAP exhibits antiviral activity against a variety of viruses such as MLV (Gao et al. 2002), alphaviruses (Bick et al. 2003), filoviruses (Muller et al. 2007), HIV-1 (Zhu et al. 2011), and hepatitis-B virus (Mao et al. 2013). Interestingly, many putative L1 restriction factors also are involved in antiviral defense (*i.e*., a subset of APOBEC3 proteins, TREX1, MOV10, SAMHD1, and RNaseL). L1 elements have been active in mammalian genomes for ~160 million years (Burton et al. 1986, Smit et al. 1995, Yang et al. 2014). Thus, it is tempting to speculate that some host factors, such as ZAP, may have first evolved to combat endogenous retrotransposons and subsequently were co-opted as viral restriction factors (Sawyer et al. 2004, Sawyer et al. 2005, Sawyer and Malik 2006, Kerns et al. 2008). Indeed, identifying host factors that modulate L1 retrotransposition may prove to be an effective strategy to identify host antiviral factors.

Methods

Cell Culture

HeLa-JVM cells were grown in high-glucose DMEM (Gibco) supplemented with 10% FBS (Gibco), 100 U/mL penicillin-streptomycin (Invitrogen), and 0.29 mg/mL L-glutamine (Gibco) (Moran et al. 1996). HeLa-HA (Hulme et al. 2007) and PA-1 (Zeuthen et al. 1980) cells were grown in MEM (Gibco) with 10% FBS, 100 U/mL penicillin-streptomycin, 0.29 mg/mL L-glutamine, and 0.1 mM nonessential amino acids (Gibco). Cell lines were maintained at 37°C with 7% CO_2 in humidified incubators (Thermo Scientific).

Plasmids

Oligonucleotide sequences and cloning strategies used in this study are available upon request. All human L1 plasmids contain the indicated fragments of L1.3 (accession no. L19088) (Sassaman et al. 1997) DNA cloned into pCEP4 (Invitrogen) unless otherwise indicated. A CMV promoter augments expression of all L1 and cDNA expressing plasmids unless noted otherwise. L1 plasmids also contain an SV40 polyadenylation signal that is located downstream of the native L1 polyadenylation signal. All plasmid DNA was prepared with a Midiprep Plasmid DNA Kit (Qiagen).

The following cDNA expression plasmids were obtained from OriGene: CDK9 (SC119344); DDX21 (SC108813); GNB2L1 (SC116322); hnRNPDL (SC107613); MOV10 (SC126015); MATR3 (SC113375); ZAP-S (ZC3HAV1 transcript variant 2) (SC101064); ZAP-S-tGFP (GFP-tagged ZC3HAV1 transcript variant 2) (RG208070); hnRNPA2B1 (SC313092); IGF2BP3 (SC111161); PURA (SC127792); UPF1 (SC118343).

The following cDNA expression plasmids were obtained from Open Biosystems: hnRNPL (6174088); LARP2 (5164712); LARP4 (5219803); SYNCRIP (5495201).

The following cDNA expression plasmids were obtained from Addgene: rZAP (pcDNA4-TO-Myc-rZAP; Addgene plasmid#: 17381, kindly provided by Dr.

Stephen Goff) (Gao et al., 2002) and ZAP-L (pcDNA4 huZAP(L); Addgene plasmid#: 45907, kindly provided by Dr. Harmit Malik) (Kerns et al. 2008).

pJM101/L1.3: is a pCEP4-based plasmid that expresses a human L1 (L1.3) equipped with an *mneoI* retrotransposition indicator cassette. L1 expression is augmented by a CMV promoter located upstream of the L1 5' UTR and an SV40 polyadenylation signal that is located downstream of the native L1 polyadenylation signal (Dombroski et al. 1993, Freeman et al. 1994, Moran et al. 1996, Sassaman et al. 1997)

pJM101/L1.3FLAG: was derived from pJM101/L1.3 and contains a single FLAG epitope on the carboxyl-terminus of ORF1p. Dr. Huira Kopera (University of Michigan Medical School) constructed the plasmid.

pAlu*neo*^{Tet}: expresses an Alu element cloned from intron 5 of the human *NF1* gene (Wallace et al. 1991) that is marked with the *neo*^{Tet} reporter gene. The reporter (Esnault et al. 2002) was subcloned upstream of the Alu poly adenosine tract (Dewannieux et al. 2003).

pCEP/GFP: is a pCEP4 based plasmid that expresses the humanized renilla green fluorescent protein (hrGFP) coding sequence from phrGFP-C (Stratagene), which is located downstream of the pCEP4 CMV promoter (Alisch et al. 2006).

pJJ101/L1.3: is a pCEP4 based plasmid that contains an active human L1 (L1.3) equipped with an *mblastI* retrotransposition indicator cassette (Kopera et al. 2011).

pJJ105/L1.3: is similar to pJJ101/L1.3, but contains a D702A missense mutation in the RT active site of L1.3 ORF2 (Kopera et al. 2011).

pJM101/L1.3Δneo: is a pCEP4 based plasmid that contains an active human L1 (L1.3) (Wei et al. 2001).

pLRE3-EF1-*mEGFP*ΔIntron: is a pBSKS-II+ based plasmid that expresses an active human L1 (LRE3) that is tagged with an *EGFP* cassette (*mEGFPI*) containing an antisense, intronless copy of the EGFP gene. A UbC promoter

drives EGFP expression. An EF1α promoter drives L1 expression (Wissing et al. 2011).

pAD2TE1: is similar to pJM101/L1.3 except that it was modified to contain a *T7 gene10* epitope-tag on the carboxyl-terminus of ORF1p and a TAP epitope-tag on the carboxyl-terminus of ORF2p. The 3′-UTR contains the *mneoI* retrotransposition indicator cassette (Doucet et al. 2010).

pJBM2TE1: is similar to pAD2TE1 except that the pCEP4 backbone was modified to contain the puromycin resistance (PURO) gene in place of the hygromycin resistance gene.

pLRE3-*mEGFPI*: is a pCEP4 based plasmid that contains an active human L1 (LRE3) equipped with an *mEGFPI* retrotransposition indicator cassette (Ostertag et al. 2000, Garcia-Perez et al. 2010). The pCEP4 backbone was modified to contain a puromycin resistance (PURO) gene in place of the hygromycin resistance gene. The CMV promoter also was deleted from the vector; thus, L1 expression is driven only by the native 5′ UTR (Ostertag et al. 2000).

pJM111-LRE3-*mEGFPI*: is identical to pLRE3-*mEGFPI* except that it contains two missense mutations in ORF1 (RR261-262AA), which render the L1 retrotransposition-defective (Moran et al. 1996). Mr. William Giblin (University of Michigan Medical School) constructed the plasmid (Zhang et al. 2014).

pG_F21: contains an 8.8 kb fragment which includes a full length mouse G_F21 L1 element that contains the *mneoI* indicator cassette (Goodier et al. 2001).

pZfL2-2: is a pCEP4 based plasmid that contains the ZfL2-2 ORF (ZL15, accession no. AB211150) cloned upstream of the *mneoI* indicator cassette (Sugano et al. 2006).

pCEP4smL1: contains a codon optimized full-length mouse element (derived from $L1_{spa}$) containing the *mneoI* indicator cassette (Han and Boeke 2004).

ZAP-S/1-311: encodes the ZAP-S amino acid sequence from 1-311 and the following sequence of non-templated amino acids (IIIYTGFLFCCGFFFFFFFLEGVSLCCPGWS).

ZAP-S/Δ72-372: was derived by deleting the *Sfo*I-*Xho*I fragment from ZC3HAV1 transcript variant 2 (OriGene, SC101064), and expresses a ZAP-S mutant protein that lacks amino acid sequence from 72-372.

ZAP-S/Δ310-645-tGFP: expresses a ZAP-S mutant protein that lacks amino acid sequence from 310-645 and contains a carboxyl terminus tGFP epitope tag.

ZAP-S/Δ72-372-tGFP: expresses a ZAP-S mutant protein that lacks amino acid sequence from 72-372 and contains a carboxyl terminus tGFP epitope tag.

LARP5: was derived by cloning LARP5 cDNA (Open Biosystems, 40118844) into pcDNA3 (Invitrogen).

LARP1: was constructed by cloning the LARP1 cDNA (Open Biosystems, 3138935) into pcDNA3 (Invitrogen).

pK_A3A: expresses HA-tagged APOBEC3A and was a generous gift from Dr. Brian Cullen (Bogerd et al. 2006).

pDCP1α-GFP: expresses a GFP-tagged version of DCP1α and was a generous gift from Dr. Gregory Hannon (Liu et al. 2005).

pG3BP-GFP: expresses a GFP-tagged version of G3BP and was a generous gift from Dr. Jamal Tazi (Tourriere et al. 2003).

pcDNA6/TR: expresses the blasticidin resistance gene and was obtained from Invitrogen.

Immunoprecipitation of L1 ORF1p

HeLa-JVM cells were seeded in T-175 flasks (BD Falcon) at ~6-8×10^6 cells/flask and transfected the next day with 20 µg of plasmid DNA using 60 µL of FuGENE HD (Promega). Approximately 48 hours post-transfection, hygromycin B (Gibco) (200 µg/mL) was added to the medium to select for transfected cells. After approximately one week of hygromycin selection, cells were washed 3 times with ice cold PBS and collected with a rubber policeman into 50 mL conical tubes (BD Falcon). Cells were then pelleted at 1,000×g and frozen at -80°C. To produce whole cell lysates (WCL), frozen cell pellets were rapidly thawed and

then lysed in ~3 mL (1 mL lysis buffer per 100 mg of cell pellet) of lysis buffer (20 mM Tris-HCl (pH 7.5), 150 mM NaCl, 10% glycerol, 1 mM EDTA, 0.1% IGEPAL CA-630 (Sigma), 1X complete EDTA-free protease inhibitor cocktail (Roche)) on ice for 30 minutes. WCLs were then centrifuged at 15,000×g for 15 minutes at 4°C. Supernatants were transferred to a clean tube and protein concentration was determined using the Bradford reagent assay (BioRad). For the IP, ~1 mL of the supernatant (~3 mg total protein) was pre-cleared with ~15 µL (packed gel volume) of mouse IgG-agarose beads (Sigma) for 4 hours at 4°C. Pre-cleared supernatants were then mixed with ~15 µL (packed gel volume) of EZview Red ANTI-FLAG M2 Affinity Gel (Sigma) and incubated overnight with rotation at 4°C. The beads then were rinsed 3x with 0.5 mL of lysis buffer, and then washed 3 times with 0.5 mL of lysis buffer for 10 minutes per wash on ice with gentle agitation. Protein complexes were eluted from the beads by adding ~70 µL of 2X NuPAGE LDS Sample Buffer (Novex), supplemented with NuPAGE Sample Reducing Agent (Novex), directly to the washed beads and incubating for 10 minutes at 70°C. Following incubation, the beads were pelleted and the sample was transferred to a fresh tube. For SDS-PAGE analysis, 20 µL of the IP were loaded onto a 4-15% gradient midi-gel (BioRad) and run under reducing conditions. Gels were silver stained using the SilverQuest Silver Staining Kit (Novex) to visualize proteins.

Protein Identification by LC-MS/MS

The Proteomics Facility at the Fred Hutchinson Cancer Research Center (Seattle, WA) conducted protein identification experiments. Excised silver-stained gel slices were destained and subjected to in-gel proteolytic digestion with trypsin as described (Shevchenko et al. 1996). Following gel-slice digestion, the digestion products were desalted using C18-micro ZipTips (Millipore) and were dried by vacuum centrifugation. The resultant peptide samples were resuspended in 7 µL of 0.1% formic acid and 5 µL were analyzed by liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS). LC-MS/MS analysis was performed using an LTQ Orbitrap XL mass spectrometer (Thermo Scientific). The LC system, configured in a vented format (Licklider et al. 2002),

consisted of a fused-silica nanospray needle (PicoTip emitter, 50 µm ID) (New Objective) packed in-house with Magic C18 AQ 100A reverse-phase medium (25 cm) (Michrom Bioresources Inc.) and a trap (IntegraFrit Capillary, 100 µm ID) (New Objective) containing Magic C18 AQ 200A reverse-phase medium (2 cm) (Michrom Bioresources Inc.). The peptide samples were loaded onto the column and chromatographic separation was performed using a two mobile-phase solvent system consisting of 0.1% formic acid in water (A) and 0.1% acetic acid in acetonitrile (B) over 60 min from 5% B to 40% B at a flow rate of 400 nL/minutes. The mass spectrometer operated in a data-dependent MS/MS mode over the *m/z* range of 400-1800. For each cycle, the five most abundant ions from each MS scan were selected for MS/MS analysis using 35% normalized collision energy. Selected ions were dynamically excluded for 45 seconds.

For data analysis, raw MS/MS data were submitted to the Computational Proteomics Analysis System (CPAS), a web-based system built on the LabKey Server v11.2 (Rauch et al. 2005) and searched using the X!Tandem search engine (Craig and Beavis 2004) against the International Protein Index (IPI) human protein database (v3.75), which included additional common contaminants such as BSA and trypsin. Search results were compared between the pJM101/L1.3FLAG lane and the pJM101/L1.3 lane to generate a list of candidate L1 ORF1p associated proteins unique to the pJM101/L1.3FLAG immunoprecipitation. The search output files were analyzed and validated by ProteinProphet (Nesvizhskii et al. 2003). Peptide hits were filtered with PeptideProphet (Keller et al. 2002) error rate ≤ 0.05, and proteins with probability scores of ≥ 0.95 were accepted. Suspected contaminants (*e.g.* keratin) were filtered from the final L1 RNP candidate list.

L1 Retrotransposition Assays

The cultured cell retrotransposition assay was carried out essentially as described (Moran et al. 1996, Wei et al. 2000). For retrotransposition assays with L1 constructs tagged with *mblastI*, HeLa-JVM cells were seeded at ~1-2×10^4 cells/well in a 6-well plate (BD Falcon). Within 24 hours, each well was

transfected with 1 µg of plasmid DNA (0.5 µg L1 plasmid + 0.5 µg cDNA plasmid or pCEP4) using 3 µL of FuGENE 6 transfection reagent (Promega). Four days post-transfection, blasticidin (EMD Millipore) containing medium (10 µg/mL) was added to cells to select for retrotransposition events. Medium was changed every two days. After ~8 days of selection, cells were washed with PBS, fixed, and then stained with crystal violet to visualize colonies. To control for transfection efficiency and off-target effects of cDNA plasmids, in parallel with retrotransposition assays, HeLa-JVM cells were plated in 6-well plates at 500-1,000 cells/well and transfected with 0.5 µg pcDNA6/TR (Invitrogen) plasmid + 0.5 µg cDNA plasmid using 3 µL of FuGENE 6 transfection reagent (Promega). The pcDNA6/TR control assays were treated with blasticidin in the same manner as for retrotransposition assays.

For retrotransposition assays with L1 constructs tagged with *mneoI*, HeLa-JVM cells were transfected as described above. Two days after transfection, cells were treated with medium supplemented with G418 (Gibco) (500 µg/mL) for ~10-12 days. As a control, HeLa cells were plated at ~2×10^4 cells/well in a 6-well plate and transfected with 0.5 µg pcDNA3 (Invitrogen) plasmid + 0.5 µg cDNA plasmid using 3 µL of FuGENE 6 transfection reagent (Promega). The pcDNA3 control assays were treated with G418 in the same manner as for retrotransposition assays.

Alu Retrotransposition Assays

For Alu retrotransposition assays (Dewannieux et al. 2003), ~4×10^5 HeLa-HA cells were plated per well of a 6-well plate (BD Falcon) and transfected with 0.67 µg of pJM101/L1.3Δneo + 0.67 µg of pAlu*neo*^{Tet} + 0.67 µg of cDNA plasmid using 6 µL FuGENE HD (Promega). Three days post-transfection, cells were grown in the presence of G418 (500µg/mL) to select for Alu retrotransposition events. As a control, HeLa-HA cells were plated at ~4×10^5 cells/well in a 6-well plate and transfected with 0.67 µg of pcDNA3 (Invitrogen) + 0.67 µg of pAlu*neo*^{Tet} + 0.67 µg of cDNA plasmid using 6 µL of FuGENE HD (Promega). The pcDNA3

control assays were treated with G418 in the same manner as for Alu retrotransposition assays.

siRNA knockdown and pLRE3-*mEGFPI* retrotransposition assays

In experiments to study the effect of endogenous proteins on L1 retrotransposition, HeLa cells (~8×10^5 cells) were plated in 60 mm tissue culture dishes (BD Falcon). The next day, the cells were transfected with 50 nM of a control siRNA pool (D-001810-10, ON-TARGETplus Non-targeting Pool, Thermo Scientific) or siRNA against ZAP (L-017449-01-0005, ON-TARGETplus Human ZC3HAV1 (56829) siRNA - SMARTpool, Thermo Scientific) or MOV10 (L-014162-00-0005, ON-TARGETplus Human MOV10 (4343) siRNA - SMARTpool, Thermo Scientific) using the DharmaFECT 1 transfection reagent (Thermo Scientific). Twenty-four hours after siRNA treatment, cells were transfected with pLRE3-*mEGFPI* or pJM111-LRE3-*mEGFPI* (5 µg), using 15 µL of FuGENE HD transfection reagent (Roche). After 48 hours, cells were trypsinized and an aliquot of the cells (~2×10^6 cells) was used to monitor endogenous protein levels (72 hours after siRNA treatment) by western blot analysis (see below for list of primary antibodies). Blots were analyzed using an Odyssey CLx (LI-COR) with the following secondary antibodies: IRDye 800CW Donkey anti-Rabbit IgG (1:10,000) (LI-COR) and IRDye 680RD Donkey anti-Mouse IgG (1:10,000) (LI-COR). Knockdown efficiencies were calculated using LI-COR Image Studio Software (v3.1.4) and are the average of three independent experiments. Endogenous tubulin was used as the normalization control. The remaining cells were re-plated at ~2×10^5 cells/well of a 6-well plate and cultured in medium supplemented with puromycin (5 µg/ml, Gibco/Life Technologies) to select for cells transfected with pLRE3-*mEGFPI*. After 4 days of puromycin selection, the percentage of GFP positive cells was determined by flow cytometry using an Accuri C6 flow cytometer (BD Biosciences).

RNP Isolation

RNPs were isolated as previously described (Kulpa and Moran 2005). Briefly, HeLa-JVM cells were seeded onto 60 mm tissue culture dishes (BD

Falcon) and 24 hours later cells were co-transfected with 2.5 µg of pJBM2TE1 and 2.5 µg of the indicated cDNA plasmid using 15 µL of FuGENE HD (Promega). Approximately two days after transfection, puromycin (5 µg/mL) was added to culture medium to select for cells transfected with pJBM2TE1. After ~3 days of puromycin selection (5 days after transfection), cells were lysed in RNP lysis buffer (150 mM NaCl, 5 mM $MgCl_2$, 20 mM Tris-HCl (pH 7.5), 10% glycerol, 1mM DTT, 0.1% NP-40, and 1x complete EDTA-free protease inhibitor cocktail (Roche)). Following lysis, whole cell lysates were centrifuged at 12,000xg for 10 minutes at 4°C, and then the cleared lysate was layered onto a sucrose cushion (8.5% and 17% sucrose) and subjected to ultracentrifugation at 4°C for 2 hours at 178,000xg. The supernatant was discarded and the resulting pellet was resuspended in water supplemented with 1x complete EDTA-free protease inhibitor cocktail (Roche). Approximately 20 µg (total protein) of the RNP sample or ~30 µg (total protein) of the cleared whole cell lysate (supernatant post 12,000xg centrifugation) were then analyzed by western blot. Blots were analyzed using an Odyssey® CLx (LI-COR) with the following secondary antibodies: IRDye 800CW Donkey anti-Rabbit IgG (1:10,000) (LI-COR) and IRDye 680RD Donkey anti-Mouse IgG (1:10,000) (LI-COR).

To simultaneously analyze the effects of ZAP-S on ORF1p and EGFP protein expression, HeLa-JVM cells were seeded onto 10 cm dishes (~2.7×10^6 cells/dish) (BD Falcon) and transfected with 10 µg of plasmid DNA (5.0 µg pLRE3-EF1A-*mEGFP*ΔIntron + 5.0 µg cDNA plasmid or pCEP4) using 30 µL of FuGENE HD. After 48 hours, cells were harvested with trypsin and then subjected to flow cytometry to isolate GFP expressing cells. Approximately 1.2-1.7×10^6 GFP positive cells were collected for each transfection condition using a MoFlo Astrios cell sorter (Beckman Coulter). The GFP gate was set using untransfected HeLa-JVM cells. The sorted cells were lysed as described in the IP procedure and lysates were then subjected to western blotting using standard procedures. For all other protein expression analyses, HeLa-JVM cells were seeded at ~4×10^5 cells/well in 6-well plates and transfected with 2 µg of plasmid DNA with 6 µL of FuGENE HD. Cells were collected 48 hours after transfection

using a rubber policeman and lysates were prepared as described above. Western blots were visualized using either the SuperSignal West Femto Chemiluminescent Substrate (Pierce) or SuperSignal West Pico Chemiluminescent Substrate (Pierce) and Hyperfilm ECL (GE Healthcare).

Northern Blots

HeLa-JVM cells were seeded in T-175 flasks (BD Falcon) and transfected with 20 µg of plasmid DNA (10 µg pJM101/L.13Δneo + 10 µg cDNA plasmid) using 60 µL FuGENE HD. Two days after transfection, cell pellets were collected and frozen at -80°C. Frozen cell pellets were then thawed and total RNA was extracted with TRIzol reagent (Ambion), and then poly(A)+ RNA was prepared from total RNA using an Oligotex mRNA kit (Qiagen). Each sample (~1.5 µg of poly(A)+ RNA) was subjected to glyoxal gel electrophoresis and northern blotting using the NorthernMax-Gly Kit (Ambion) according to the manufacturer's protocol. Following electrophoresis, RNA was transferred to BrightStar Nylon membranes (Invitrogen) and then cross-linked using UV light. For northern blot detection, membranes were prehybridized for ~4 hours at 68°C in NorthernMax Prehybridization/Hybridization Buffer (Ambion), and then incubated with a strand specific RNA probe (final concentration of probe ~3×10^6 cpm ml^{-1}) overnight at 68°C. For band quantification, northern blot films were analyzed using ImageJ software (Schneider et al. 2012).

Strand-specific RNA probes were generated using the MAXIscript T3 system (Invitrogen). The 5UTR99 (Belancio et al. 2006) probe corresponds to bases 7-99 of the L1.3 5' UTR and the ORF2_5804 probe corresponds to nucleotides 5560-5804 of the L1.3 sequence. RNA probe templates for T3 reactions were generated by PCR using pJM101/L1.3Δneo as a PCR template with the following primer pairs:

(5UTR99: 5'-GGAGCCAAGATGGCCGAATAGGAACAGCT-3' and 5'-AATTAACCCTCAAAGGGACCTCAGATGGAAATGCAG-3');

(ORF2_5804: 5'- GACACATGCACACGTATGTTTATT-3' and 5'-AATTAACCCTCACTAAAGGGTGAGTGAGAATATGCGGTGTTT-3').

The T3 promoter sequence (underlined) was added to the reverse primer of each primer pair. The pTRI-β-actin-125-Human Antisense Control Template (Applied Biosystems) was used in T3 reactions as a template to generate the β-actin RNA probe. Each northern blot experiment was independently repeated three times with similar results.

Immunofluorescence Microscopy

Immunofluorescence microscopy was performed essentially as described (Doucet et al. 2010) with modifications. Briefly, cells were plated on round glass cover slips (Fisher) in a 12-well plate or into 4-well chambered glass slides (Fisher) and transfected ~24 hours later with 0.5 µg of plasmid DNA using 1.5 µL of FuGENE 6 transfection reagent. To visualize proteins, approximately 48 hours post-transfection cells were washed with 1x PBS, fixed with 4% paraformaldehyde for 10 minutes and then treated with ice-cold methanol for 1 minute. Next, cells were incubated for 30 minutes at 37°C in 1x PBS + 3% BSA. Cells then were incubated with primary antibodies in 1x PBS + 3% BSA for 1 hour at 37°C. Cells were washed three times with 1x PBS (10 minutes per wash) and then incubated with appropriate, fluorescently-labeled secondary antibodies diluted in 1x PBS for 30 minutes at 37°C. The following secondary antibodies were used for indirect immunofluorescence: Alexa Fluor 488 conjugated Goat anti-Mouse and Goat anti-Rabbit (Invitrogen) (1:1000), Alexa Fluor 546 conjugated Goat anti-Mouse and Goat anti-Rabbit IgG (Invitrogen) (1:1000), and Cy5 conjugated Donkey anti-Rabbit IgG (H+L) (Jackson ImmunoResearch) (1:100). To obtain images, a cover slip and/or slide was visually scanned and representative images were captured using a Leica SP5X confocal microscope (63x/1.4 objective; section thickness 1 µm).

Combined RNA FISH (Fluorescence *in situ* Hybridization) and immunofluorescence

Cells were plated on round glass cover slips (Fisher) in a 12-well plate and transfected ~24 hours later with 0.5 µg of plasmid DNA using 1.5 µL of FuGENE 6 transfection reagent. Approximately 48 hours after transfection, cells were fixed

with 4% paraformaldehyde for 10 minutes and then permeabilized with 0.2% Triton X-100 in 1x PBS for 7 minutes. Following permeabilization, coverslips were incubated for 5 minutes in FISH (fluorescence *in situ* hybridization) wash buffer (2x SSC, 10% formamide) for 5 minutes. To visualize L1 RNA, coverslips were then incubated with 300 nM FISH probes (sequences below) in FISH hybridization buffer (2x SSC, 10% formamide, 1% dextran sulphate) for ~4 hours at 37°C. Following hybridization, cells were incubated for 30 minutes in FISH wash buffer at 37°C and then incubated with FISH wash buffer + 3% BSA for an additional 30 minutes at 37°C. To visualize L1 ORF1p by immunofluorescence, coverslips then were incubated with αORF1p antibodies (1:2000) in 1x PBS + 3% BSA for 1 hour at 37°C. Cells were washed three times with 1x PBS (10 minutes per wash). Cells were incubated with Alexa Fluor 546 conjugated Goat anti-Rabbit IgG (Invitrogen) (1:1000) in 1x PBS + DAPI (50 ng/mL) for 30 minutes at 37°C. Coverslips were mounted on slides with VECTASHIELD mounting media (Vector Laboratories). Combined RNA FISH/immunofluorescence samples were imaged with a Zeiss Axioplan2 microscope (63x objective; Axiovision 4.8 software). RNA FISH/immunofluorescence images (Figures 3.6A-3.6D) were globally processed using the Photoshop CS6 (version 13.0 x64) Levels tool to adjust input levels. The L1 RNA was labeled using 21 Quasar670-labelled anti-sense oligonucleotide probes complimentary to sequences within the L1.3 5' UTR (probes were designed and produced by Biosearch Technologies, Petaluma, CA). The sequences of the 21 L1 probes are as follows: 5'-aaatcaccgtcttctgcgtc-3', 5'-ggtacctcagatggaaatgc-3', 5'-cactccctagtgagatgaac-3', 5'-ccctttctttgactcagaaa-3', 5'-aatattcgggtgggagtgac-3', 5'-cttaagccggtctgaaaagc-3', 5'-caggtgtgggatatagtctc-3', 5'-tgctagcaatcagcgagatt-3', 5'-ttgcagtttgatctcagact-3', 5'-tttgtttacctaagcaagcc-3', 5'-cagaggtggagcctacagag-3', 5'-ctgtctttttgtttgtctgt-3', 5'-cacttaagtctgcagaggtt-3', 5'-ctctcttcaaagctgtcaga-3', 5'-ttgaggaggcagtctgtctg-3', 5'-ctgcaggtctgttggaatac-3', 5'-ttctaacagacaggaccctc-3', 5'-cctttctggttgttagtttt-3', 5'-gatgggttttcggtgtagat-3', 5'-gtctttgatgatggtgatgt-3', 5'-tttgtggttttatctacttt-3'.

Primary Antibodies

Polyclonal antibodies against peptide sequences 31-49 of L1.3 ORF1p (αORF1p) were raised in rabbits and affinity-purified (Open Biosystems). αCDK9 (2316), αUPF1 (9435), and αGFP (2955) were obtained from Cell Signaling Technology. αhnRNPL (NBP1-67852), αILF3 (EPR3627), αLARP1 (NBP1-19128), αMATR3 (NB100-1761), αNCL (NB100-1920SS), and αDHX9 (NB110-40579) were obtained from Novus Biologicals. αFAM120A (ab83909), αPURA (ab79936), and αHA tag (ab9110) were obtained from Abcam. αMOV10 (SAB1100141), αZAP (Anti-ZC3HAV1 (HPA047818)), and αTubulin (T9026) were obtained from Sigma. αZC3HAV1 (16820-1-AP) was obtained from Proteintech. αeIF3 (p110) (sc-28858) was obtained from Santa Cruz Biotechnology. αT7-Tag mouse monoclonal (69522-3) was obtained from Novagen. αTAP rabbit polyclonal (CAB1001) was obtained from Thermo Scientific.

Acknowledgements

We thank Darko Plavcic for assistance with plasmid DNA preparation and retrotransposition assays. We thank Dr. Huira Kopera, Dr. Jose L. Garcia-Perez, Dr. JoAnn Sekiguchi, Dr. Anthony Antonellis, and J.V.M. lab members for critical reading of this manuscript and helpful discussions during the course of this study. We thank Dr. Phil Gafken and the Fred Hutchinson Cancer Research Center Proteomics Core for their invaluable help with protein mass spectrometry experiments. We thank the flow cytometry and DNA sequencing core facilities at the University of Michigan. We thank Bruce Donohoe and the Microscopy & Image Analysis Core at the University of Michigan for assistance with immunofluorescence microscopy. We are grateful to Nancy Leff and Dr. Stephanie Bolz for invaluable editorial assistance during the preparation of this manuscript.

Figure 3.1: The identification of host proteins immunoprecipitated with L1 ORF1p-FLAG.

(A) Schematic of L1 constructs: pJM101/L1.3 expresses a human L1 (L1.3) (Sassaman et al. 1997) containing an *mneoI* retrotransposition indicator cassette within the L1 3' UTR. The pJM101/L1.3FLAG construct is identical to pJM101/L1.3, but contains a single FLAG epitope on the carboxyl-terminus of ORF1p. Both constructs were cloned into a pCEP4 mammalian expression vector. A CMV promoter augments L1 expression and an SV40 polyadenylation signal (pA) is located downstream of the native L1 polyadenylation signal. *(B) Results of immunoprecipitation experiments:* Whole cell lysates from HeLa cells transfected with either pJM101/L1.3 or pJM101/L1.3FLAG were subjected to immunoprecipitation using an anti-FLAG antibody. The proteins then were separated by SDS-PAGE, visualized by silver staining, and subjected to LC-MS/MS. An ~40 kDa band corresponding to the theoretical molecular weight of ORF1p is visible in the pJM101/L1.3FLAG lane (*). Black bars indicate the approximate molecular weights of the ORF1p-FLAG interacting proteins. Molecular weight standards (kDa) are shown on the left hand side of the gel. *(C) Validation of the ORF1p-FLAG immunoprecipitation:* Western blot experiments using an antibody specific to amino acids 31-49 of L1.3 ORF1p verified the enrichment of ORF1p-FLAG in pJM101/L1.3FLAG, but not pJM101/L1.3 immunoprecipitation reactions. Cells transfected with the pCEP4 vector served as a negative control. *(D) Validation of putative ORF1p-FLAG interacting proteins:* Western blot images of the pJM101/L1.3FLAG and pJM101/L1.3 immunoprecipitation (IP) reactions. The pCEP4 lanes denote whole cell lysates derived from HeLa cells transfected with an empty pCEP4 vector (~1.0% input). Primary antibodies used to probe western blots are indicated to the left of the images. Immunoprecipitation reactions were conducted in either the absence (left) or presence (right) of RNaseA (10 µg/mL). The putative cellular functions of the ORF1p-FLAG interacting proteins are indicated on the right hand side of the blots.

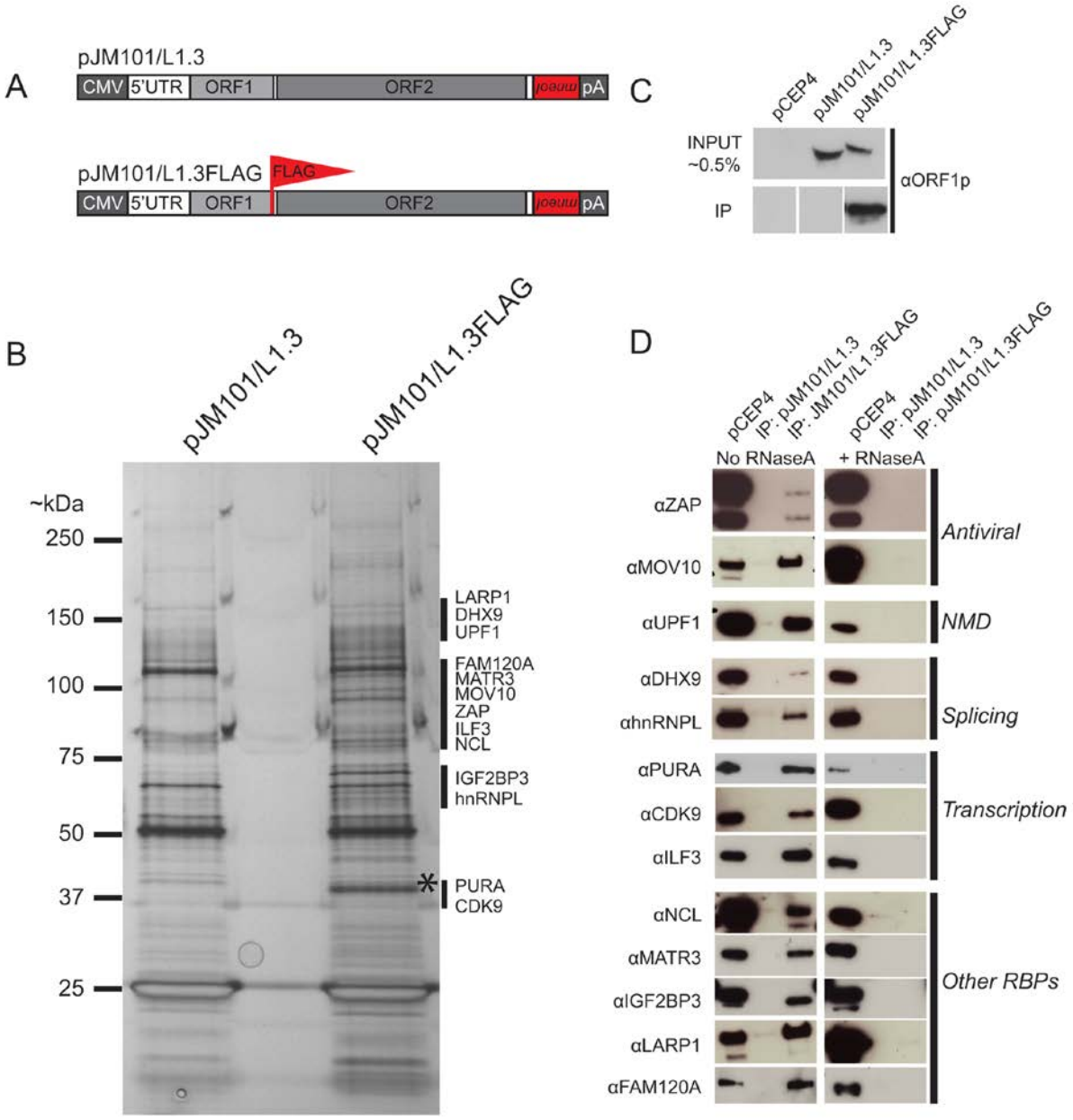


Figure 3.1: The identification of host proteins immunoprecipitated with L1 ORF1p-FLAG.

Figure 3.2: Several of the ORF1p-FLAG interacting proteins inhibit L1 retrotransposition.

(A) Schematic of the cultured-cell retrotransposition assay: HeLa cells were transfected with an engineered human L1.3 construct (pJJ101/L1.3) marked with a blasticidin indicator cassette (*mblastI*). The pJJ101/L1.3 construct was cloned into a pCEP4 mammalian expression vector. A CMV promoter augments L1 expression and an SV40 polyadenylation signal (pA) is located downstream of the native L1 polyadenylation signal. The *mblastI* cassette is cloned into the L1 3' UTR antisense to the L1 and contains a blasticidin deaminase gene that is disrupted by an intron in the L1 sense orientation. The blasticidin deaminase gene can only be expressed when the L1 transcript is spliced, reverse transcribed, and inserted into genomic DNA (Moran et al. 1996, Wei et al. 2000). *(B) Schematic of the pJJ101/L1.3 retrotransposition screen*: To analyze the effect of the ORF1p-FLAG interacting proteins on L1 retrotransposition, HeLa cells were co-transfected with equal amounts of pJJ101/L1.3 and a cDNA plasmid expressing one of the candidate ORF1p-FLAG interacting proteins or a pCEP4 empty vector. To control for potential off-target effects, HeLa cells also were co-transfected with a control plasmid (pcDNA6/TR) that expresses the blasticidin deaminase gene and a cDNA plasmid expressing one of the candidate proteins or a pCEP4 empty vector. Both assays were subjected to the same blasticidin selection regimen. The resultant number of blasticidin-resistant colonies in pcDNA6/TR control assays provides a visual, quantitative readout of the effect of cDNA overexpression on the ability of cells to grow in the presence of blasticidin. *(C) Results of pJJ101/L1.3 retrotransposition screen*: HeLa cells were co-transfected with pJJ101/L1.3 and each of the indicated cDNA expressing plasmids. L1 retrotransposition was assayed in 6-well tissue culture plates. The X-axis indicates the cDNA that was co-transfected with pJJ101/L1.3. The bracketed number next to each cDNA indicates the number of independent experiments. The Y-axis indicates L1 retrotransposition activity after accounting for cDNA toxicity (see Figure 3.2B). Retrotransposition activity (black bars) is normalized to the pCEP4 empty vector control. Error bars represent the standard deviation for each set of experiments. The red dotted line indicates a 50% inhibition of retrotransposition activity.

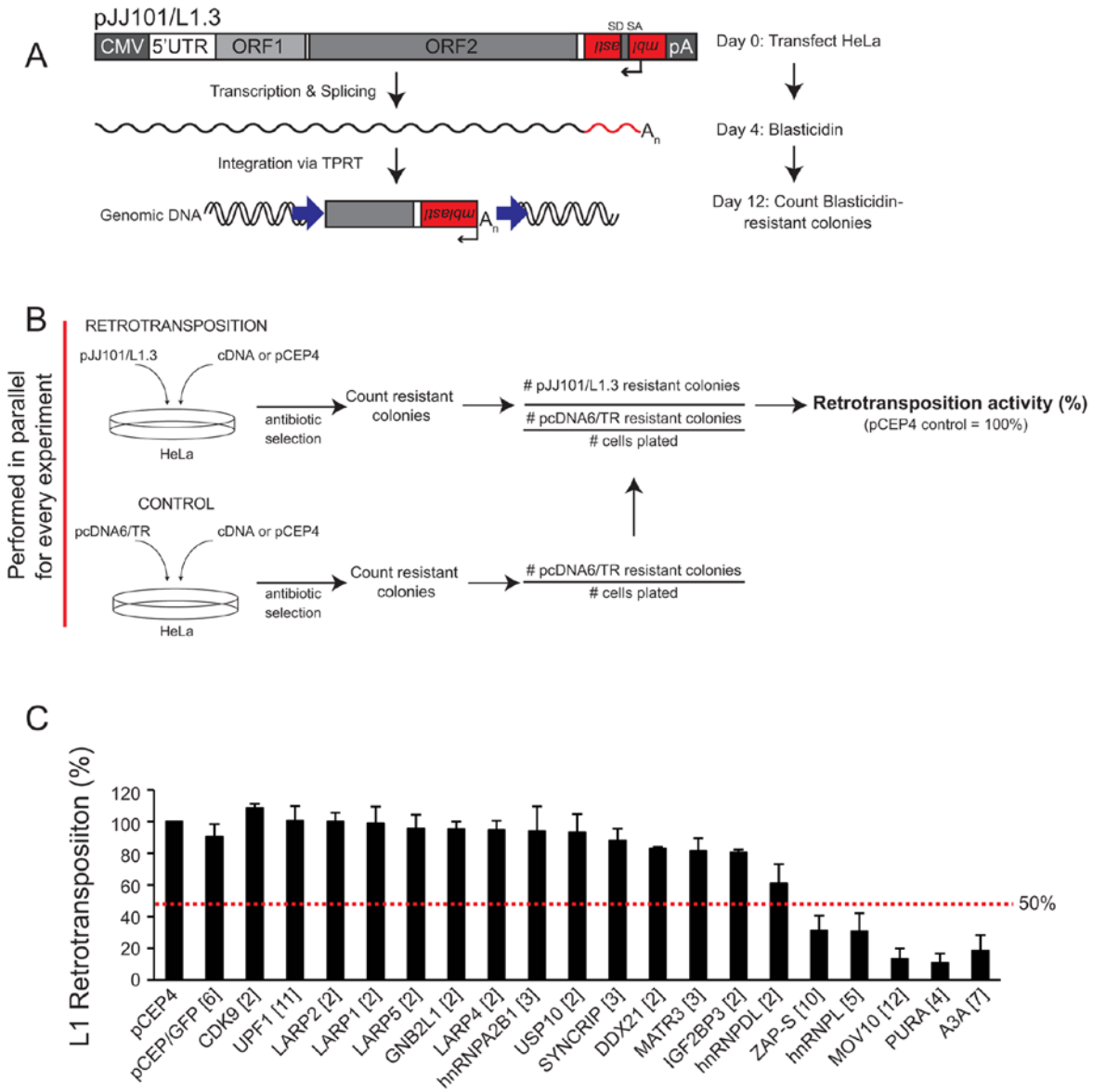


Figure 3.2: Several of the ORF1p-FLAG interacting proteins inhibit L1 retrotransposition.

Figure 3.3: ZAP-S inhibits LINE and Alu retrotransposition.

(A) ZAP inhibits L1 retrotransposition: Top panel: Schematics of ZAP constructs. Depicted are the relative positions of the zinc-finger domains (light gray rectangles), cysteine-histidine (CCCH) zinc-fingers (vertical black bars), and PARP-like domain (dark gray rectangles) of the ZAP-L and ZAP-S expression constructs. ZAP-L contains a carboxyl-terminal HA tag (blue rectangle labeled HA). The ZAP-S/1-311 construct contains an additional 31 amino acids at the carboxyl terminus. The ZAP-S/ 372 harbors a deletion that removes the CCCH zinc fingers (See Methods). Middle panel: Results of the retrotransposition assays. The X-axis indicates the cDNA co-transfected with pJJ101/L1.3 or pcDNA6/TR. The Y-axis indicates pJJ101/L1.3 retrotransposition activity (black bars), or pcDNA6/TR colony formation activity (white bars). All values have been normalized to the pCEP4 empty vector control (100%). The numbers above the bar graphs indicate the number of independent experiments performed with each cDNA expression construct. Error bars represent standard deviations. Bottom panel: A single well of a representative six-well tissue culture plate, displaying blasticidin-resistant colonies from the pJJ101/L1.3 retrotransposition assay (top, black rectangle) and the pcDNA6/TR control assay (bottom, white rectangle). *(B) ZAP inhibits Alu retrotransposition:* The X-axis indicates the cDNA co-transfected with pJM101/L1.3Δneo and pAlu*neo*^{Tet}. The Y-axis indicates the retrotransposition efficiency. All values are normalized to the pCEP4 empty vector control (100%). Control assays using a plasmid that expresses the neomycin phosphotransferase gene (pcDNA3) were conducted similarly to pcDNA6/TR control assays as outlined in Figure 3.2B. Representative images of G418-resistant HeLa foci from the Alu retrotransposition assay are shown below the bar graph. The results are the average of three independent experiments. Error bars indicate standard deviations. *(C) ZAP inhibits the retrotransposition of mouse and zebrafish LINE elements.* The X-axis indicates the cDNA that was co-transfected with human L1 (pJM101/L1.3 (black bars)), mouse L1 (pG_F21 (dark grey bars)), zebrafish L2 (pZfL2-2 (light grey bars)), or synthetic mouse L1 (pCEPsmL1 (white bars)). The Y-axis indicates the retrotransposition efficiency. Representative images of G418-resistant HeLa cell foci are shown below the bar graph. Control assays using a plasmid that expresses the neomycin phosphotransferase gene (pcDNA3) were conducted similarly to pcDNA6/TR control assays outlined in Figure 3.2B. All values are normalized to the pCEP4 empty vector control (100%). Error bars indicate standard deviations. *(D) The depletion of ZAP enhances L1 retrotransposition:* Top panels: Western blots of whole cell lysates derived from mock HeLa cell transfections or HeLa cells transfected with indicated siRNAs. Blue arrows point to the approximate location of ZAP-L and ZAP-S. Bottom panel: The bar graph depicts pLRE-*mEGFP1* retrotransposition activity following siRNA treatment. The X-axis indicates the siRNA. The Y-axis indicates the pLRE-*mEGFP1* retrotransposition efficiency normalized to the control siRNA (set to 1). Retrotransposition efficiency values are reported as the mean from four independent experiments. Error bars indicate the standard deviations. Asterisks indicate statistically significant differences from the control siRNA experiments (two-tailed t test/$p<0.05$).

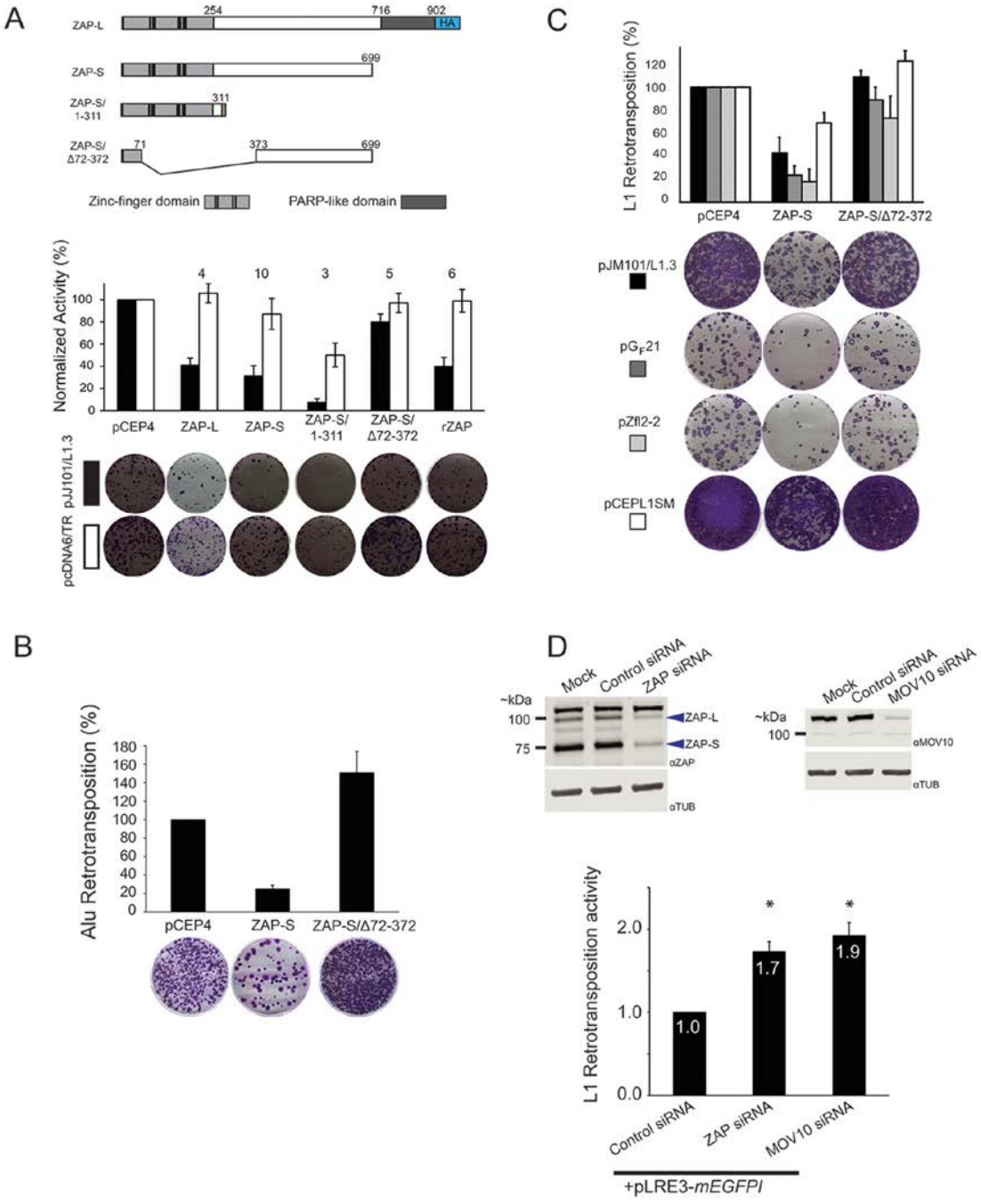


Figure 3.3: ZAP-S inhibits LINE and Alu retrotransposition.

Figure 3.4: The effect of ZAP-S on L1 RNA and L1 protein expression.

(A) Schematic of pJM101/L1.3Δneo: Bold black lines indicate the approximate location of probes (5UTR99 and ORF2_5804) used in the northern blot experiments. pJM101/L1.3Δneo is expressed from a pCEP4 vector. A CMV promoter augments L1 expression and an SV40 polyadenylation signal (pA) is located downstream of the native L1 polyadenylation signal. *(B) Results of northern blots:* Top panel: HeLa cells were co-transfected with pJM101/L1.3Δneo and either the indicated ZAP-S expression plasmids or an empty pCEP4 vector. Northern blot images depict the effect of ZAP-S overexpression on polyadenylated L1 RNA levels. The constructs transfected into HeLa cells are indicated above each lane. UTF indicates untransfected HeLa cells and serves as a negative control. Probes (5UTR99 and ORF2_5804) are indicated in the top left corner of the respective blots. The black arrow indicates the position of the full-length L1 RNA. The blue and yellow arrows indicate shorter L1 RNA species. The experiment was repeated three times with similar results. Actin served as a loading control. RNA size standards (~kb) are shown at the right of the blot image. Bottom panel: Quantification of northern blot bands. The X-axis indicates the cDNA expression construct that was co-transfected with pJM101/L1.3Δneo. The Y-axis indicates relative band intensity normalized to pCEP4 controls (100%). Black bars represent the full-length L1 band. Blue and yellow bars represent the smaller L1 RNA bands, corresponding to the colored arrows, respectively, in the top panel. The results are the average of three independent experiments. Error bars represent standard deviations. *(C) Schematic of pJBM2TE1*: The construct contains a T7 epitope tag on the carboxyl-terminus of ORF1p and a TAP tag on the carboxyl-terminus of ORF2p. An *mneoI* retrotransposition indicator cassette is present in the 3' UTR. pJMB2TE1 is expressed from a pCEP4 backbone, which has been modified to contain a puromycin selectable marker. A CMV promoter augments L1 expression and an SV40 polyadenylation signal (pA) is located downstream of the native L1 polyadenylation signal. *(D) ZAP-S decreases the accumulation of the L1-encoded proteins*: HeLa cells were co-transfected with pJBM2TE1 and the plasmids indicated above each lane. UTF indicates untransfected HeLa cells and serves as a negative control. Depicted are western blots using whole cell lysates (WCL, top panel) or RNP fractions (RNP, bottom panel). Blue arrows indicate the positions of ORF2p, ORF1p, ZAP-S, and ZAP-S/ -372. The eIF3 protein is used as a loading control. Representative images are shown. The experiments were repeated three times with similar results.

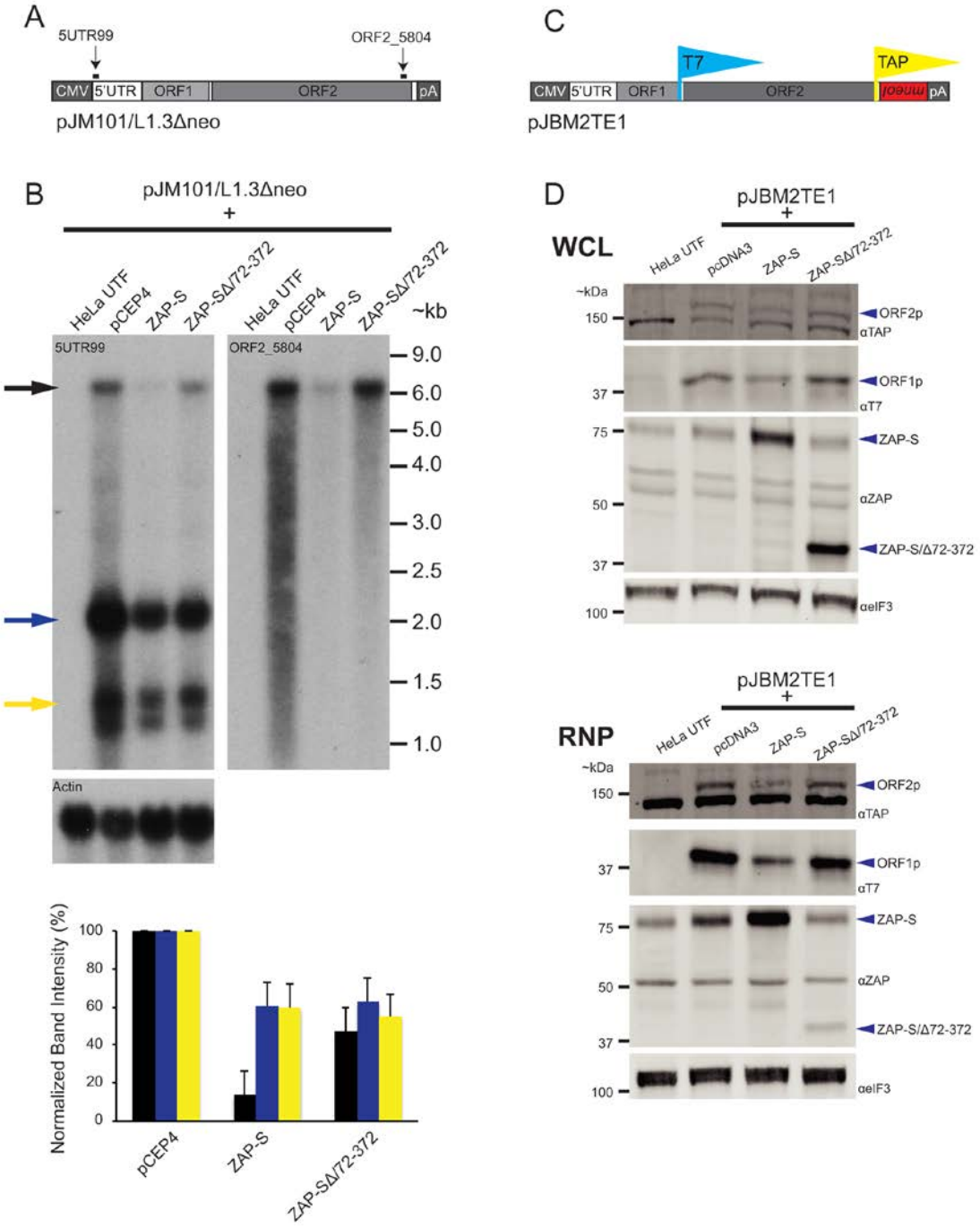


Figure 3.4: The effect of ZAP-S on L1 RNA and L1 protein expression.

Figure 3.5: The co-localization of ORF1p and ZAP in cytoplasmic foci.

(A) Co-localization of transfected ORF1p and ZAP-S in HeLa cells: ORF1p (red) expressed from pJM101/L1.3Δneo co-localizes with ZAP-S-tGFP (green). The experiment was repeated five times with similar results. *(B) Co-localization of transfected ORF1p with endogenous ZAP in cytoplasmic foci in HeLa cells:* ORF1p-T7 (green) expressed from pAD2TE1 co-localizes with endogenous ZAP (red). The experiment was repeated five times with similar results. *(C) Co-localization of transfected ZAP-S-tGFP with endogenous ORF1p in cytoplasmic foci in PA-1 cells*: ZAP-S-tGFP (green) co-localizes with endogenous ORF1p (red). PA-1 experiments were repeated twice with similar results. *(D-E) The ZAP-S zinc-finger domain is necessary for co-localization with ORF1p in HeLa cells:* ORF1p (red) expressed from pJM101/L1.3Δneo co-localizes with ZAP-S/Δ310-645-tGFP (green) (panel D). ORF1p (red) expressed from pJM101/L1.3Δneo forms cytoplasmic foci that do not contain ZAP-S/Δ72-372-tGFP (green) (panel E). The right-most image of each panel represents a merged image. The cell type is indicated at the top left (yellow), the protein name is listed on the bottom left, and the name of the primary antibody used (*italicized*) is annotated at the bottom right. Nuclei were stained with DAPI (blue) and the scale bar represents 25 μM.

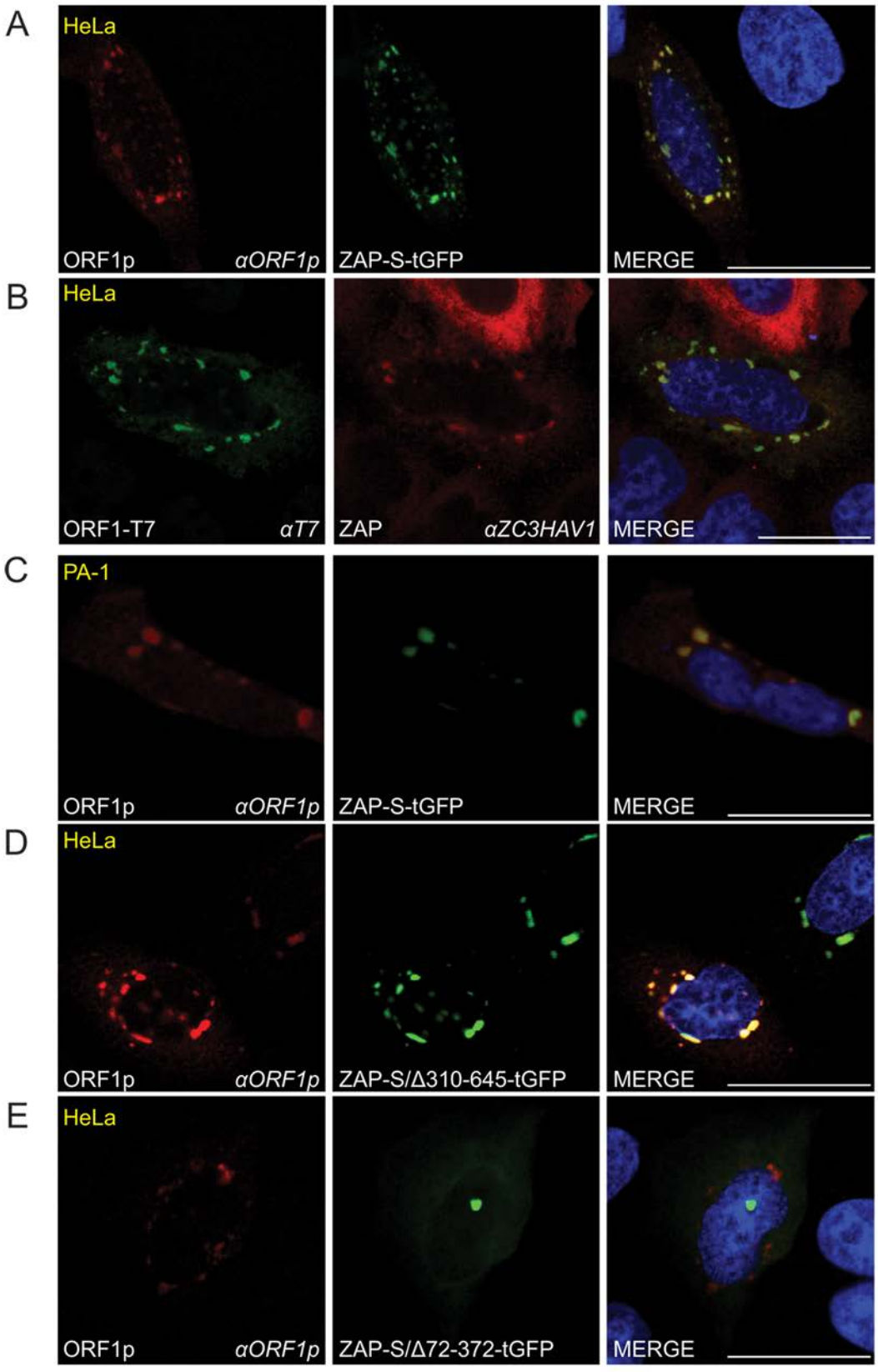

A
HeLa
ORF1p
αORF1p
ZAP-S-tGFP
MERGE
B
HeLa
ORF1-T7
αT7
ZAP
αZC3HAV1
MERGE
C
PA-1
ORF1p
αORF1p
ZAP-S-tGFP
MERGE
D
HeLa
ORF1p
αORF1p
ZAP-S/Δ310-645-tGFP
MERGE
E
HeLa
ORF1p
αORF1p
ZAP-S/Δ72-372-tGFP
MERGE

Figure 3.6: The co-localization of ZAP with L1 RNA and ORF1p in HeLa cells.

(A) Co-localization of transfected L1 RNA and ORF1p: ORF1p (red) expressed from pJM101/L1.3 co-localizes with L1 RNA (magenta). *(B) Co-localization of transfected L1 RNA and ORF1p with transfected ZAP-S-tGFP in cytoplasmic foci:* ORF1p (red) and L1 RNA (magenta) expressed from pJM101/L1.3 co-localize with ZAP-S-tGFP (green). *(C-D) The ZAP-S zinc-finger domain is necessary for co-localization with ORF1p:* ORF1p (red) and L1 RNA (magenta) expressed from pJM101/L1.3 co-localize with ZAP-S/Δ310-645-tGFP (green) (panel C). ZAP-S/Δ72-372-tGFP (green) diffusely distributes throughout the cytoplasm, while ORF1p (red) expressed from pJM101/L1.3 forms cytoplasmic foci with L1 RNA (magenta) (panel D). The right-most image of each panel represents a merged image. The name of the protein or RNA is indicated at the bottom left, and the name of the primary antibody used (*italicized*) is annotated at the bottom right of each image. Nuclei were stained with DAPI (blue) and the scale bar represents 25 μM. Experiments were repeated three times with similar results.

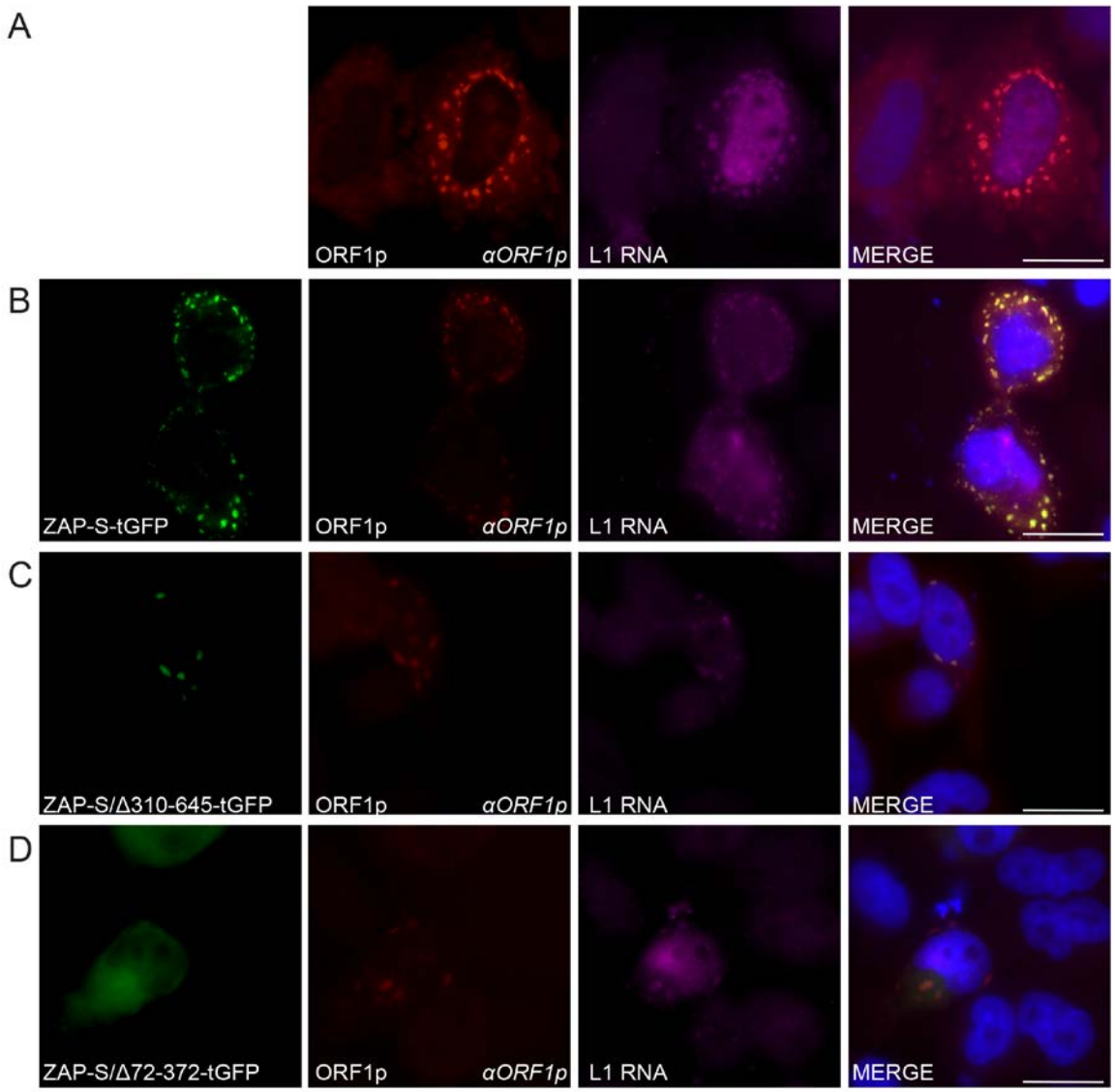


Figure 3.6: The co-localization of ZAP with L1 RNA and ORF1p in HeLa cells.

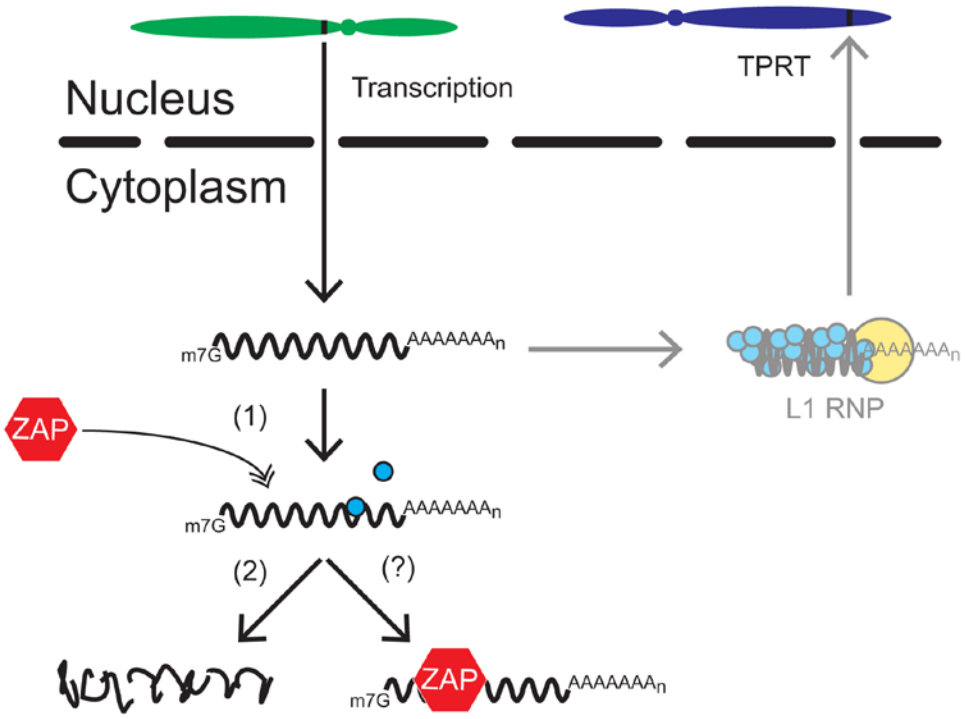


Figure 3.7: A working model for how ZAP restricts L1 retrotransposition.

Once a genomic L1 (black rectangle located on the green chromosome) is transcribed, the resultant bicistronic L1 mRNA is exported to the cytoplasm for translation. L1 ORF1p (blue circles) and ORF2p (yellow circle) bind back to L1 mRNA to form an L1 RNP. The L1 RNP gains access to the nucleus where a new L1 copy is inserted into genomic DNA by the process of TPRT (black rectangle located on the blue chromosome). In ZAP-mediated restriction, ZAP (red hexagon) interacts with L1 mRNA in the cytoplasm (1), which we propose leads to the destabilization of L1 RNA (2) and/or a block in translation (?) through the recruitment of other cellular factors (*e.g.*, SG associated proteins, RNA decay proteins) involved in RNA metabolism.

Figure 3.8: Supporting data for Figure 3.1.

(A) The FLAG epitope on ORF1p is compatible with retrotransposition: Constructs were tested in a transient HeLa cell retrotransposition assay (Moran et al. 1996, Wei et al. 2000). The X-axis indicates the L1 plasmid. The Y-axis indicates the retrotransposition efficiency. Retrotransposition assays were normalized to pJM101/L1.3 (100%). The pJM105/L1.3 plasmid serves as a negative control and harbors a point mutation in the ORF2p RT domain that renders the element inactive (Moran et al. 1996). Representative results from a single experiment are depicted below the graph. The assay was repeated two times with similar results. *(B) Immunoprecipitation reactions conducted using various wash conditions:* Top panel: HeLa cells were transfected with pCEP4, pJM101/L1.3, or pJM101/L1.3FLAG and were subjected to lysis using two different salt concentrations (500 mM NaCl (left gel) or 150 mM NaCl (right gel)). Shown are the images of silver stained gels from immunoprecipitation reactions. The black rectangles indicate the cropped image depicted in Figure 3.1B. Molecular weight standards (~kDa) are shown on the left side of the gel. Bottom panel: Image of full western blot used in Figure 3.1C. The black rectangle indicates the cropped lanes depicted in Figure 3.1C. Molecular weight standards (~kDa) are shown on the left side of the gel. *(C) Immunoprecipitation reactions under different lysis buffer conditions:* Silver stained gels of IP fractions from untransfected HeLa (HeLa UTF) or HeLa cells transfected with pJM101/L1.3FLAG. Lysis buffer contained either 0.1% CHAPS (left gel) or 1.0% Triton X-100 (right gel). Black arrows correspond to the approximate location of ORF1p-FLAG; black bars indicate the approximate location of proteins enriched in the pJM101/L1.3FLAG lane. Molecular weight standards (kDa) are shown on the left side of the gels.

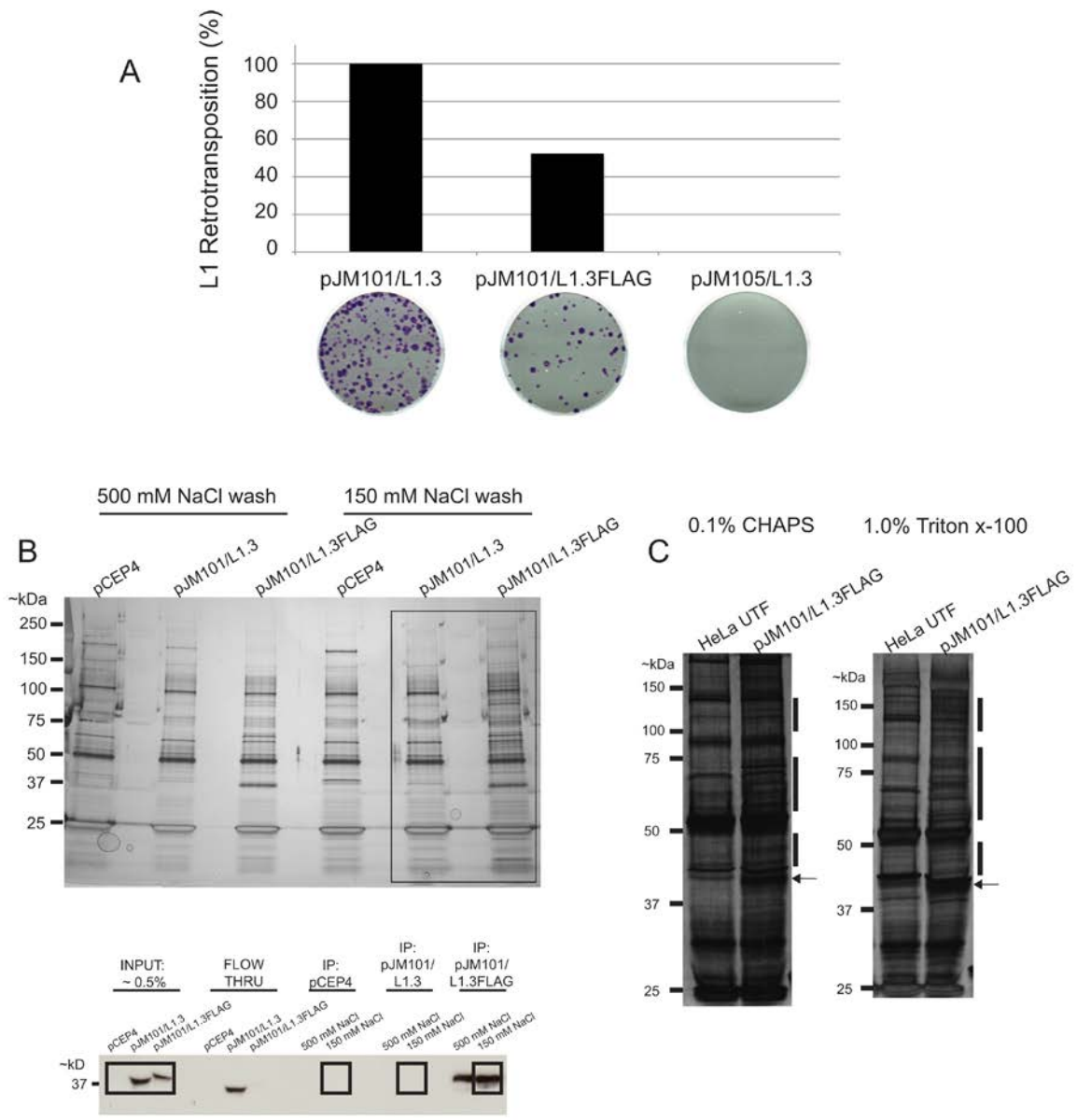


Figure 3.8: Supporting data for Figure 1.

Figure 3.9: Supporting data for Figure 3.3.

(A-C) Transfected ZAP is expressed in HeLa cells: Western blots of whole cell lysates demonstrate the expression of ZAP-L (panel A) and ZAP-S and ZAP-S/□72-372 (panels B and C) 48 hours post-transfection. UTF indicates untransfected HeLa cells. The antibodies are indicated at the right side of the blots. Blue arrows indicate the approximate locations of the ZAP proteins. Tubulin serves as a loading control. Molecular weight standards (kDa) are shown on the left side of the blots. *(D) The depletion of endogenous ZAP enhances L1 retrotransposition:* Flow cytometry was used to determine the percentage of EGFP-positive, live-gated cells for each siRNA transfection condition (noted above the plots). The X-axis depicts the scattering at 533 nm; the Y-axis depicts the scattering at 585 nm. The EGFP-positive gate was set using the retrotransposition-deficient negative control, pJM111-LRE3-*mEGFPI* (Moran et al. 1996, Zhang et al. 2014).

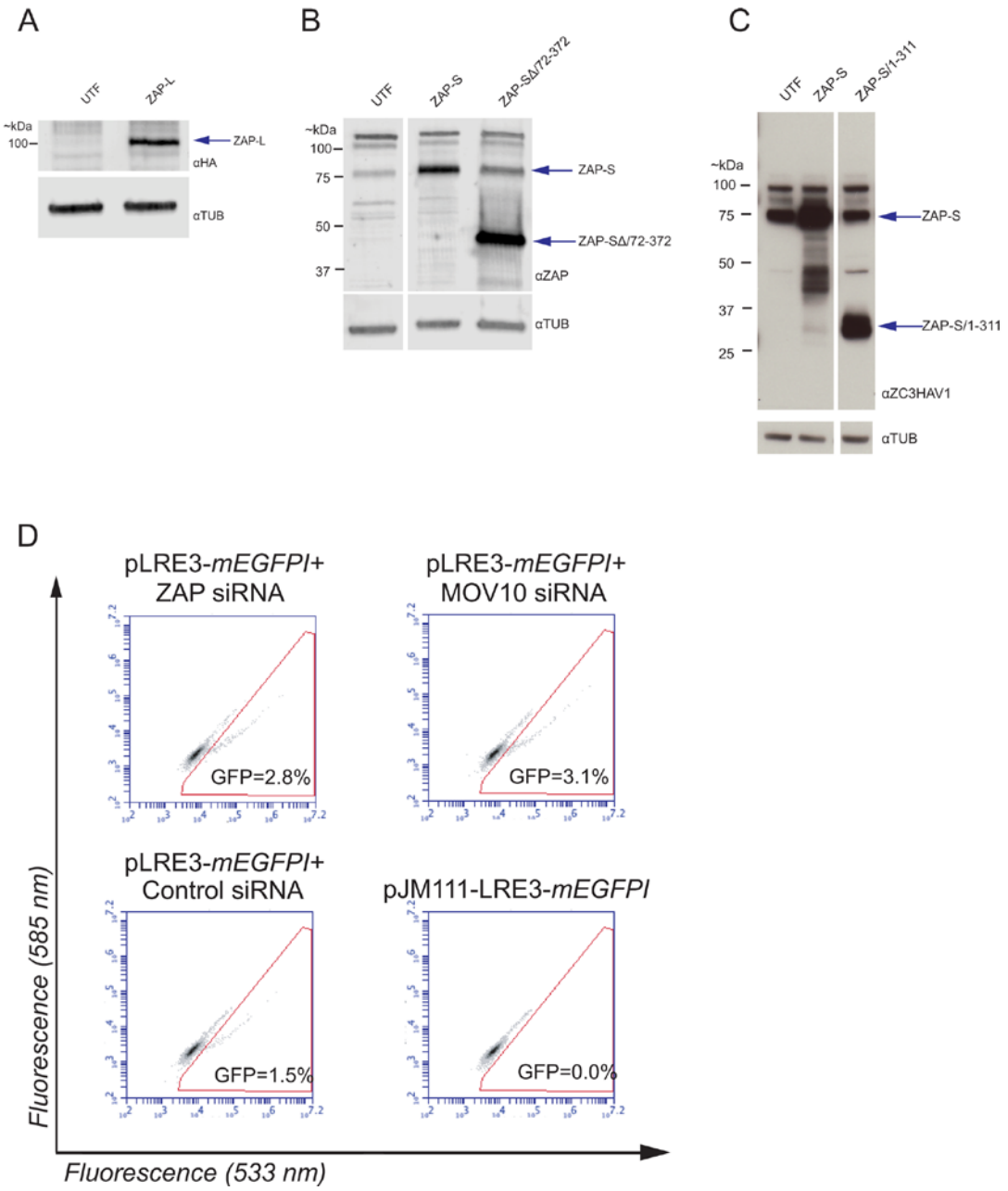


Figure 3.9: Supporting data for Figure 3.3.

Figure 3.10: Supporting data for Figure 3.4 ZAP-S preferentially suppresses the expression of ORF1p.

(A) A schematic of the pLRE3-EF1-mEGFPΔIntron: pLRE3-EF1-*mEGFP*ΔIntron expresses a human L1 (LRE3) that is tagged with an *mEGFPI* expression cassette that lacks an intron. The human elongation factor-1 alpha (EF1α) promoter (arrow) augments L1 transcription. The ubiquitin C (UbC) promoter (upside down arrow) drives EGFP transcription. *(B) ZAP-S inhibits ORF1p expression:* Western blots were conducted using whole cell lysates derived from cells co-transfected with pLRE3-EF1-*mEGFP*ΔIntron and the ZAP-S expression plasmid or pCEP4 indicated above each lane. UTF indicates whole cell lysates from untransfected HeLa cell. Antibodies are indicated on the right side of each blot. Tubulin is used as a loading control. Western blot images depict a representative experiment that was repeated three times with similar results. Notably, upon extended exposure times ORF1p was able to be visualized in the ZAP-S lane. *(C) ZAP-S does not inhibit EFGP expression and/or accumulation:* HeLa cells were co-transfected with pLRE3-EF1-*mEGFP*ΔIntron and the indicated expression plasmids (noted above the plots). Flow cytometry was used to determine the percentage of EGFP-positive, live-gated cells for each condition. UTF indicates untransfected HeLa cells. The EGFP-positive gate was set using the UTF sample as a negative control. The X-axis depicts the percentage of EGFP positive cells. The Y-axis indicates the side scattering profile (SSC). Approximately 1.2 - 1.7 x 10^6 GFP positive cells were collected and analyzed for each transfection condition.

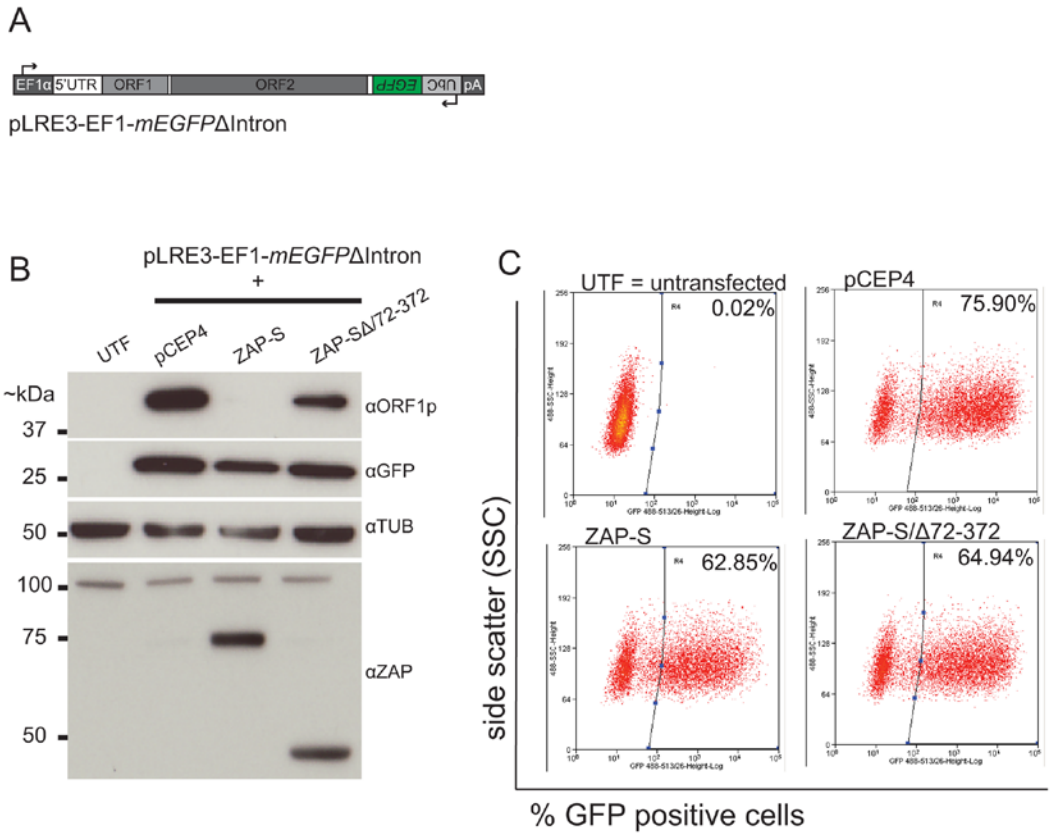
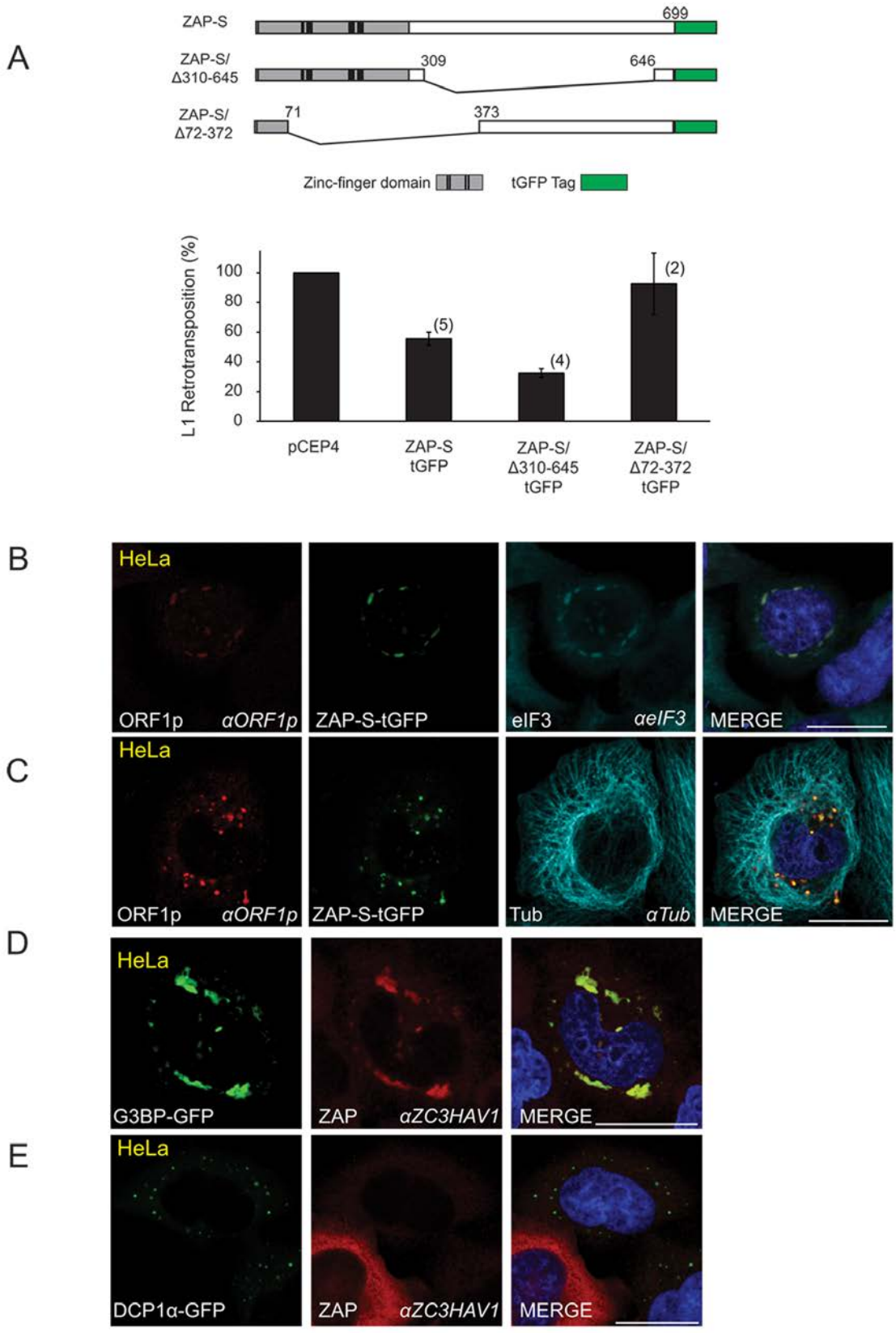


Figure 3.10: Supporting data for Figure 3.4 ZAP-S preferentially suppresses the expression of ORF1p.

Figure 3.11: Supporting data for Figure 3.5.

(A) ZAP-S-tGFP inhibits retrotransposition in HeLa cells: Top panel: Schematics of tGFP-tagged ZAP constructs. Depicted are the relative positions of the zinc-finger domains (light gray rectangles), cysteine-histidine (CCCH) zinc-fingers (vertical black bars), and tGFP tag (green rectangles) ZAP-S expression constructs. Bottom panel: Results of pJJ101/L1.3 retrotransposition assays. The X-axis indicates the cDNA co-transfected with pJJ101/L1.3 or pcDNA6/TR. The Y-axis indicates pJJ101/L1.3 retrotransposition activity (black bars). All values have been normalized to the pCEP4 empty vector control (100%). The numbers above the bar graphs indicate the number of biological replicates performed with each cDNA expression construct. Error bars represent standard deviations. *(B-E) ORF1p and ZAP co-localize with stress granules in HeLa cells:* HeLa cells were co-transfected with pJM101/L1.3Δneo and ZAP-S-tGFP; proteins were visualized by direct immunofluorescence. ORF1p co-localizes with ectopic ZAP-S-tGFP and eIF3 in cytoplasmic foci (panel B). ORF1p co-localizes with ectopic ZAP-S-tGFP (panels A and B), but not with tubulin (panel C). GFP-tagged G3BP co-localizes with endogenous ZAP in cytoplasmic foci (panel D). GFP-tagged DCP1 □ forms cytoplasmic punctate structures, which do not appear to co-localize with endogenous ZAP (panel E). The right-most image in each panel represents a merged image. The cell type is indicated at the top left (yellow), the protein name is listed on the bottom left, and the name of the primary antibody used (*italicized*) is annotated at the bottom right. Nuclei were stained with DAPI (blue) and the scale bar represents 25 μM.


A
ZAP-S
699
ZAP-S/
Δ310-645
309
646
ZAP-S/
Δ72-372
71
373
Zinc-finger domain
tGFP Tag
L1 Retrotransposition (%)
100
80
60
40
20
0
(5)
(4)
(2)
pCEP4
ZAP-S
tGFP
ZAP-S/
Δ310-645
tGFP
ZAP-S/
Δ72-372
tGFP
B
HeLa
ORF1p
αORF1p
ZAP-S-tGFP
eIF3
αeIF3
MERGE
C
HeLa
ORF1p
αORF1p
ZAP-S-tGFP
Tub
αTub
MERGE
D
HeLa
G3BP-GFP
ZAP
αZC3HAV1
MERGE
E
HeLa
DCP1α-GFP
ZAP
αZC3HAV1
MERGE

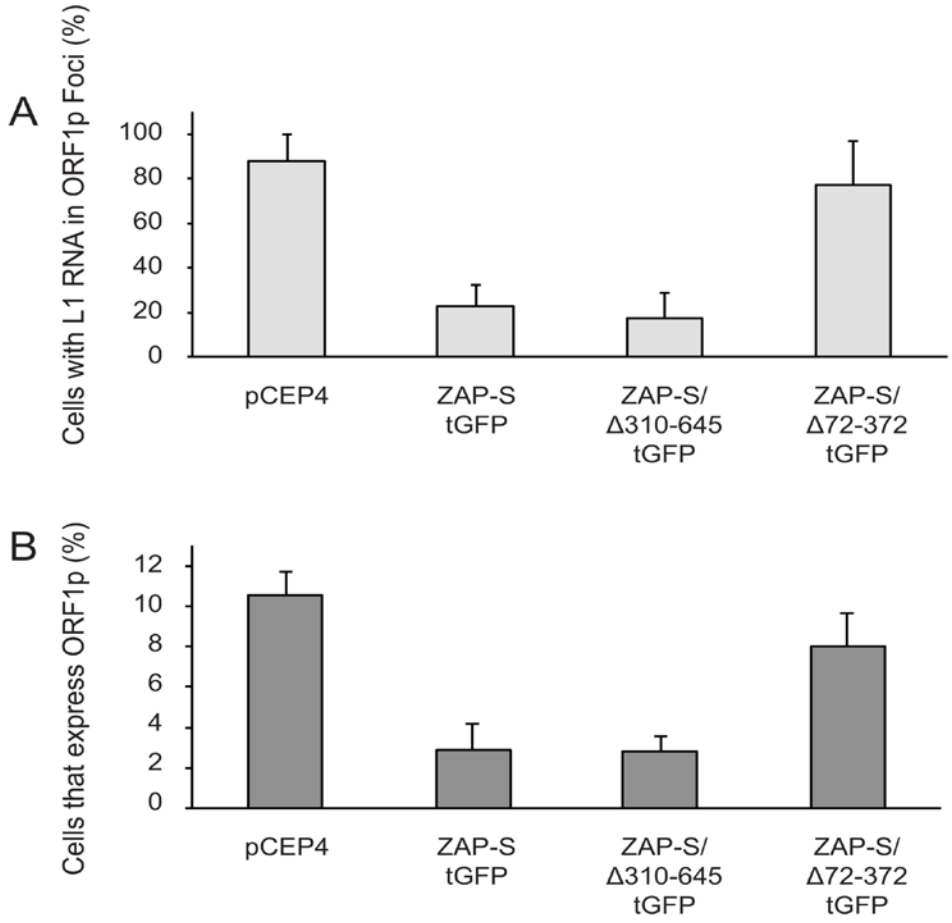


Figure 3.12: Supporting data for Figure 3.6.

(A) Fluorescence microscopy was used to determine the percentage of cells that contained L1 RNA in cytoplasmic ORF1p foci. The X-axis indicates the plasmid that was co-transfected with pJM101/L1.3. The Y-axis of the graph depicts the percentage of cells where L1 RNA was detected in cytoplasmic ORF1p foci. Experiments were repeated three times. A total of ~60 visual fields (~1600 cells) were examined amongst all three experiments and ~33-41 ORF1p foci containing cells were evaluated for each experimental condition. Error bars represent standard deviations. *(B)* Confocal microscopy was used to determine the number of ORF1p-expressing cells ~48 hours post transfection. The X-axis indicates the plasmid that was co-transfected with pJM101/L1.3Δneo. The Y-axis of the graph depicts the percentage of cells that express ORF1p. Experiments were repeated twice. Each experiment contained two biological replicates and ~1100-1500 cells were enumerated amongst all experiments for each condition. Error bars indicate standard deviations.

1. Protein Name	2. Mass (~Da)	3. Total Peptides	4. Unique Peptides	5. AACoverage	6. Goodier *et al.*	7. Taylor *et al.*	8. Peddigari *et al.*
IGF2BP3	63705	12	10	26.0%	Y	n.s.	-
MOV10	107209	12	10	10.3%	Y	n.s.	-
LARP1	123510	10	7	9.6%	Y	n.s.	-
UPF1	123036	6	6	7.9%	-	Y	-
CDK9	42777	6	5	16.4%	-	-	-
ILF3	140957	5	5	6.9%	Y	n.s.	-
NCL	95808	6	5	13.0%	Y	n.s.	Y
DHX9	76614	6	5	10.1%	Y	n.s.	-
PURA	87344	5	4	8.4%	Y	Y	-
MATR3	80552	4	4	6.1%	Y	n.s.	-
FAM120A	88359	5	4	5.7%	Y	n.s.	-
hnRNPL	73115	5	3	6.1%	Y	n.s.	Y
ZAP (ZC3HAV1)	77903	2	2	7.4%	-	n.s.	-
PAR-4	36567	4	3	16.5%	-	n.s.	-
SYNCRIP	62528	8	8	19.0%	Y	-	Y
TROVE2	59269	5	5	11.6%	Y	Y	-
MPO	73854	5	5	7.8%	-	-	-
DDX21	121888	4	4	8.8%	Y	n.s.	-
hnRNPA2B1	36006	5	4	18.3%	Y	n.s.	-
LARP5	34911	6	4	21.1%	-	n.s.	-
KIF5B	109685	4	4	5.0%	-	n.s.	-
KHSRP	90933	4	3	8.4%	-	n.s.	-
DHX15	50561	4	3	14.7%	-	n.s.	-
hnRPDL	46437	3	3	13.1%	-	n.s.	-
SFRS1	106048	2	2	3.9%	Y	n.s.	-
IGF2BP2	61640	2	2	8.0%	Y	n.s.	-
hnRNPC	35077	2	2	13.7%	Y	n.s.	-
SRRM2	33670	3	2	16.8%	-	n.s.	-
SFRS7	30841	2	2	9.2%	-	n.s.	-
ATXN2	61842	4	2	5.1%	-	n.s.	-
FUBP3	91839	2	2	4.9%	-	n.s.	-
GNB2L1	11932	2	2	18.4%	-	n.s.	-
hnRPA0	27745	5	2	16.5%	-	-	-
USP10	15763	2	2	25.5%	-	-	-
NCBP2	299615	3	2	2.3%	-	n.s.	-
NCBP1	87533	2	2	4.4%	-	n.s.	-
ANP32A	24068	3	2	17.2%	-	-	-
C21orf70	25456	2	2	11.2%	-	-	-
HIST1H1D	22350	5	2	14.9%	-	n.s.	-

Table 3.1: L1 ORF1p-interacting protein candidates identified by LC-MS/MS.

ORF1p-interacting proteins were selected based on the criteria that the protein was unique to the pJM101/L1.3FLAG IP and was identified by two or more unique peptides (peptide error ≤ 0.05; protein probability ≥ 0.95). Column 1 = protein name. Column 2 = protein mass. Column 3 = total number of identified peptides. Column 4 = number of unique peptides. Column 5 = the percentage of amino acid coverage for each of the respective proteins. Columns 6-8 = whether the proteins were identified in the indicated studies (Goodier et al. 2013, Peddigari et al. 2013, Taylor et al. 2013). Green highlighting indicates ORF1p-interacting candidates that were verified by western blot (Figure 3.1D). "Y" in columns 6-8 indicate that the protein was identified as a significant L1-interacting protein by statistical and/or direct biochemical methods; “n.s.” in column 7 indicates that the protein was identified in the study, but that it did not reach the significance threshold set by the authors.

References

Alisch, R. S., J. L. Garcia-Perez, A. R. Muotri, F. H. Gage and J. V. Moran (2006). "Unconventional translation of mammalian LINE-1 retrotransposons." Genes Dev **20**(2): 210-224.

Aravin, A. A., R. Sachidanandam, A. Girard, K. Fejes-Toth and G. J. Hannon (2007). "Developmentally regulated piRNA clusters implicate MILI in transposon control." Science **316**(5825): 744-747.

Arjan-Odedra, S., C. M. Swanson, N. M. Sherer, S. M. Wolinsky and M. H. Malim (2012). "Endogenous MOV10 inhibits the retrotransposition of endogenous retroelements but not the replication of exogenous retroviruses." Retrovirology **9**: 53.

Ashburner, M., C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin and G. Sherlock (2000). "Gene ontology: tool for the unification of biology. The Gene Ontology Consortium." Nat Genet **25**(1): 25-29.

Athanikar, J. N., R. M. Badge and J. V. Moran (2004). "A YY1-binding site is required for accurate human LINE-1 transcription initiation." Nucleic Acids Res **32**(13): 3846-3855.

Baltz, A. G., M. Munschauer, B. Schwanhausser, A. Vasile, Y. Murakawa, M. Schueler, N. Youngs, D. Penfold-Brown, K. Drew, M. Milek, E. Wyler, R. Bonneau, M. Selbach, C. Dieterich and M. Landthaler (2012). "The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts." Mol Cell **46**(5): 674-690.

Beck, C. R., P. Collier, C. Macfarlane, M. Malig, J. M. Kidd, E. E. Eichler, R. M. Badge and J. V. Moran (2010). "LINE-1 retrotransposition activity in human genomes." Cell **141**(7): 1159-1170.

Beck, C. R., J. L. Garcia-Perez, R. M. Badge and J. V. Moran (2011). "LINE-1 elements in structural variation and disease." Annu Rev Genomics Hum Genet **12**: 187-215.

Becker, K. G., G. D. Swergold, K. Ozato and R. E. Thayer (1993). "Binding of the ubiquitous nuclear transcription factor YY1 to a cis regulatory sequence in the human LINE-1 transposable element." Hum Mol Genet **2**(10): 1697-1702.

Belancio, V. P., D. J. Hedges and P. Deininger (2006). "LINE-1 RNA splicing and influences on mammalian gene expression." Nucleic Acids Research **34**(5): 1512-1521.

Belancio, V. P., A. M. Roy-Engel and P. Deininger (2008). "The impact of multiple splice sites in human L1 elements." Gene **411**(1,Äì2): 38-45.

Belancio, V. P., A. M. Roy-Engel, R. R. Pochampally and P. Deininger (2010). "Somatic expression of LINE-1 elements in human tissues." Nucleic Acids Res **38**(12): 3909-3922.

Bergemann, A. D., Z. W. Ma and E. M. Johnson (1992). "Sequence of cDNA comprising the human pur gene and sequence-specific single-stranded-DNA-binding properties of the encoded protein." Mol Cell Biol **12**(12): 5673-5682.

Bick, M. J., J.-W. N. Carroll, G. Gao, S. P. Goff, C. M. Rice and M. R. MacDonald (2003). "Expression of the Zinc-Finger Antiviral Protein Inhibits Alphavirus Replication." Journal of Virology **77**(21): 11555-11562.

Bick, M. J., J. W. Carroll, G. Gao, S. P. Goff, C. M. Rice and M. R. MacDonald (2003). "Expression of the zinc-finger antiviral protein inhibits alphavirus replication." J Virol **77**(21): 11555-11562.

Bogerd, H. P., H. L. Wiegand, B. P. Doehle, K. K. Lueders and B. R. Cullen (2006). "APOBEC3A and APOBEC3B are potent inhibitors of LTR-retrotransposon function in human cells." Nucleic Acids Res **34**(1): 89-95.

Bogerd, H. P., H. L. Wiegand, A. E. Hulme, J. L. Garcia-Perez, K. S. O'Shea, J. V. Moran and B. R. Cullen (2006). "Cellular inhibitors of long interspersed element 1 and Alu retrotransposition." Proc Natl Acad Sci U S A **103**(23): 8780-8785.

Bourc'his, D. and T. H. Bestor (2004). "Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L." Nature **431**(7004): 96-99.

Brouha, B., J. Schustak, R. M. Badge, S. Lutz-Prigge, A. H. Farley, J. V. Moran and H. H. Kazazian, Jr. (2003). "Hot L1s account for the bulk of retrotransposition in the human population." Proc Natl Acad Sci U S A **100**(9): 5280-5285.

Buchan, J. R. and R. Parker (2009). "Eukaryotic Stress Granules: The Ins and Outs of Translation." Molecular Cell **36**(6): 932-941.

Burdick, R., J. L. Smith, C. Chaipan, Y. Friew, J. Chen, N. J. Venkatachari, K. A. Delviks-Frankenberry, W. S. Hu and V. K. Pathak (2010). "P body-associated protein Mov10 inhibits HIV-1 replication at multiple stages." J Virol **84**(19): 10241-10253.

Burton, F. H., D. D. Loeb, C. F. Voliva, S. L. Martin, M. H. Edgell and C. A. Hutchison, 3rd (1986). "Conservation throughout mammalia and extensive protein-encoding capacity of the highly repeated DNA long interspersed sequence one." J Mol Biol **187**(2): 291-304.

Buzdin, A., S. Ustyugova, E. Gogvadze, T. Vinogradova, Y. Lebedev and E. Sverdlov (2002). "A new family of chimeric retrotranscripts formed by a full copy of U6 small nuclear RNA fused to the 3' terminus of l1." Genomics **80**(4): 402-406.

Castello, A., B. Fischer, K. Eichelbaum, R. Horos, B. M. Beckmann, C. Strein, N. E. Davey, D. T. Humphreys, T. Preiss, L. M. Steinmetz, J. Krijgsveld and M. W. Hentze (2012). "Insights into RNA biology from an atlas of mammalian mRNA-binding proteins." Cell **149**(6): 1393-1406.

Chen, G., X. Guo, F. Lv, Y. Xu and G. Gao (2008). "p72 DEAD box RNA helicase is required for optimal function of the zinc-finger antiviral protein." Proc Natl Acad Sci U S A **105**(11): 4352-4357.

Chen, H., C. E. Lilley, Q. Yu, D. V. Lee, J. Chou, I. Narvaiza, N. R. Landau and M. D. Weitzman (2006). "APOBEC3A is a potent inhibitor of adeno-associated virus and retrotransposons." Curr Biol **16**(5): 480-485.

Chen, S., Y. Xu, K. Zhang, X. Wang, J. Sun, G. Gao and Y. Liu (2012). "Structure of N-terminal domain of ZAP indicates how a zinc-finger protein recognizes complex RNA." Nat Struct Mol Biol **19**(4): 430-435.

Cordaux, R. and M. A. Batzer (2009). "The impact of retrotransposons on human genome evolution." Nat Rev Genet **10**(10): 691-703.

Cost, G. J. and J. D. Boeke (1998). "Targeting of human retrotransposon integration is directed by the specificity of the L1 endonuclease for regions of unusual DNA structure." Biochemistry **37**(51): 18081-18093.

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

Craig, R. and R. C. Beavis (2004). "TANDEM: matching proteins with tandem mass spectra." Bioinformatics **20**(9): 1466-1467.

deHaro, D., K. J. Kines, M. Sokolowski, R. T. Dauchy, V. A. Streva, S. M. Hill, J. P. Hanifin, G. C. Brainard, D. E. Blask and V. P. Belancio (2014). "Regulation of L1 expression and retrotransposition by melatonin and its receptor: implications for cancer risk associated with light exposure at night." Nucleic Acids Research **42**(12): 7694-7707.

Dewannieux, M., C. Esnault and T. Heidmann (2003). "LINE-mediated retrotransposition of marked Alu sequences." Nat Genet **35**(1): 41-48.

Dmitriev, S. E., D. E. Andreev, I. M. Terenin, I. A. Olovnikov, V. S. Prassolov, W. C. Merrick and I. N. Shatsky (2007). "Efficient translation initiation directed by the 900-nucleotide-long and GC-rich 5' untranslated region of the human retrotransposon LINE-1 mRNA is strictly cap dependent rather than internal ribosome entry site mediated." Mol Cell Biol **27**(13): 4685-4697.

Dombroski, B. A., Q. Feng, S. L. Mathias, D. M. Sassaman, A. F. Scott, H. H. Kazazian, Jr. and J. D. Boeke (1994). "An in vivo assay for the reverse transcriptase of human retrotransposon L1 in Saccharomyces cerevisiae." Mol Cell Biol **14**(7): 4485-4492.

Dombroski, B. A., S. L. Mathias, E. Nanthakumar, A. F. Scott and H. H. Kazazian, Jr. (1991). "Isolation of an active human transposable element." Science **254**(5039): 1805-1808.

Dombroski, B. A., A. F. Scott and H. H. Kazazian, Jr. (1993). "Two additional potential retrotransposons isolated from a human L1 subfamily that contains an active retrotransposable element." Proc Natl Acad Sci U S A **90**(14): 6513-6517.

Doucet, A. J., A. E. Hulme, E. Sahinovic, D. A. Kulpa, J. B. Moldovan, H. C. Kopera, J. N. Athanikar, M. Hasnaoui, A. Bucheton, J. V. Moran and N. Gilbert (2010). "Characterization of LINE-1 ribonucleoprotein particles." PLoS Genet **6**(10): e1001150.

Ergun, S., C. Buschmann, J. Heukeshoven, K. Dammann, F. Schnieders, H. Lauke, F. Chalajour, N. Kilic, W. H. Stratling and G. G. Schumann (2004). "Cell type-specific expression of LINE-1 open reading frames 1 and 2 in fetal and adult human tissues." J Biol Chem **279**(26): 27753-27763.

Esnault, C., J. F. Casella and T. Heidmann (2002). "A Tetrahymena thermophila ribozyme-based indicator gene to detect transposition of marked retroelements in mammalian cells." Nucleic Acids Res **30**(11): e49.

Esnault, C., J. Maestre and T. Heidmann (2000). "Human LINE retrotransposons generate processed pseudogenes." Nat Genet **24**(4): 363-367.

Feng, Q., J. V. Moran, H. H. Kazazian, Jr. and J. D. Boeke (1996). "Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition." Cell **87**(5): 905-916.

Freeman, J. D., N. L. Goodchild and D. L. Mager (1994). "A modified indicator gene for selection of retrotransposition events in mammalian cells." Biotechniques **17**(1): 46, 48-49, 52.

Gao, G., X. Guo and S. P. Goff (2002). "Inhibition of retroviral RNA production by ZAP, a CCCH-type zinc finger protein." Science **297**(5587): 1703-1706.

Garcia-Perez, J. L., A. J. Doucet, A. Bucheton, J. V. Moran and N. Gilbert (2007). "Distinct mechanisms for trans-mediated mobilization of cellular RNAs by the LINE-1 reverse transcriptase." Genome Res **17**(5): 602-611.

Garcia-Perez, J. L., M. Morell, J. O. Scheys, D. A. Kulpa, S. Morell, C. C. Carter, G. D. Hammer, K. L. Collins, K. S. O/'Shea, P. Menendez and J. V. Moran (2010). "Epigenetic silencing of engineered L1 retrotransposition events in human embryonic carcinoma cells." Nature **466**(7307): 769-773.

Gilbert, N., S. Lutz, T. A. Morrish and J. V. Moran (2005). "Multiple fates of l1 retrotransposition intermediates in cultured human cells." Mol Cell Biol **25**(17): 7780-7795.

Goodier, J. L., L. E. Cheung and H. H. Kazazian, Jr. (2012). "MOV10 RNA helicase is a potent inhibitor of retrotransposition in cells." PLoS Genet **8**(10): e1002941.

Goodier, J. L., L. E. Cheung and H. H. Kazazian, Jr. (2013). "Mapping the LINE1 ORF1 protein interactome reveals associated inhibitors of human retrotransposition." Nucleic Acids Res, 2013 **41**(15): 7401-7419.

Goodier, J. L., P. K. Mandal, L. Zhang and H. H. Kazazian, Jr. (2010). "Discrete subcellular partitioning of human retrotransposon RNAs despite a common mechanism of genome insertion." Hum Mol Genet **19**(9): 1712-1725.

Goodier, J. L., E. M. Ostertag, K. Du and H. H. Kazazian, Jr. (2001). "A novel active L1 retrotransposon subfamily in the mouse." Genome Res **11**(10): 1677-1685.

Goodier, J. L., L. Zhang, M. R. Vetter and H. H. Kazazian, Jr. (2007). "LINE-1 ORF1 protein localizes in stress granules with other RNA-binding proteins, including components of RNA interference RNA-induced silencing complex." Mol Cell Biol **27**(18): 6469-6483.

Grana, X., A. De Luca, N. Sang, Y. Fu, P. P. Claudio, J. Rosenblatt, D. O. Morgan and A. Giordano (1994). "PITALRE, a nuclear CDC2-related protein kinase that phosphorylates the retinoblastoma protein in vitro." Proc Natl Acad Sci U S A **91**(9): 3834-3838.

Grimaldi, G., J. Skowronski and M. F. Singer (1984). "Defining the beginning and end of KpnI family segments." EMBO J **3**(8): 1753-1759.

Guo, X., J.-W. N. Carroll, M. R. MacDonald, S. P. Goff and G. Gao (2004). "The Zinc Finger Antiviral Protein Directly Binds to Specific Viral mRNAs through the CCCH Zinc Finger Motifs." Journal of Virology **78**(23): 12781-12787.

Guo, X., J. Ma, J. Sun and G. Gao (2007). "The zinc-finger antiviral protein recruits the RNA processing exosome to degrade the target mRNA." Proc Natl Acad Sci U S A **104**(1): 151-156.

Han, J. S. and J. D. Boeke (2004). "A highly active synthetic mammalian retrotransposon." Nature **429**(6989): 314-318.

Hancks, D. C., J. L. Goodier, P. K. Mandal, L. E. Cheung and H. H. Kazazian, Jr. (2011). "Retrotransposition of marked SVA elements by human L1s in cultured cells." Hum Mol Genet **20**(17): 3386-3400.

Hancks, D. C. and H. H. Kazazian, Jr. (2012). "Active human retrotransposons: variation and disease." Curr Opin Genet Dev **22**(3): 191-203.

Hohjoh, H. and M. F. Singer (1996). "Cytoplasmic ribonucleoprotein complexes containing human LINE-1 protein and RNA." EMBO J **15**(3): 630-639.

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

Hohjoh, H. and M. F. Singer (1997). "Sequence-specific single-strand RNA binding protein encoded by the human LINE-1 retrotransposon." EMBO J **16**(19): 6034-6043.

Holmes, S. E., B. A. Dombroski, C. M. Krebs, C. D. Boehm and H. H. Kazazian, Jr. (1994). "A new retrotransposable human L1 element from the LRE2 locus on chromosome 1q produces a chimaeric insertion." Nat Genet **7**(2): 143-148.

Holmes, S. E., M. F. Singer and G. D. Swergold (1992). "Studies on p40, the leucine zipper motif-containing protein encoded by the first open reading frame of an active human LINE-1 transposable element." J Biol Chem **267**(28): 19765-19768.

Hui, J., K. Stangl, W. S. Lane and A. Bindereif (2003). "HnRNP L stimulates splicing of the eNOS gene by binding to variable-length CA repeats." Nat Struct Biol **10**(1): 33-37.

Hulme, A. E., H. P. Bogerd, B. R. Cullen and J. V. Moran (2007). "Selective inhibition of Alu retrotransposition by APOBEC3G." Gene **390**(1-2): 199-205.

Kao, P. N., L. Chen, G. Brock, J. Ng, J. Kenny, A. J. Smith and B. Corthesy (1994). "Cloning and expression of cyclosporin A- and FK506-sensitive nuclear factor of activated T-cells: NF45 and NF90." J Biol Chem **269**(32): 20691-20699.

Kazazian, H. H., Jr., C. Wong, H. Youssoufian, A. F. Scott, D. G. Phillips and S. E. Antonarakis (1988). "Haemophilia A resulting from de novo insertion of L1 sequences represents a novel mechanism for mutation in man." Nature **332**(6160): 164-166.

Keller, A., A. I. Nesvizhskii, E. Kolker and R. Aebersold (2002). "Empirical Statistical Model To Estimate the Accuracy of Peptide Identifications Made by MS/MS and Database Search." Analytical Chemistry **74**(20): 5383-5392.

Kernan, M. J., M. I. Kuroda, R. Kreber, B. S. Baker and B. Ganetzky (1991). "napts, a mutation affecting sodium channel activity in Drosophila, is an allele of mle, a regulator of X chromosome transcription." Cell **66**(5): 949-959.

Kerns, J. A., M. Emerman and H. S. Malik (2008). "Positive selection and increased antiviral activity associated with the PARP-containing isoform of human zinc-finger antiviral protein." PLoS Genet **4**(1): e21.

Khazina, E. and O. Weichenrieder (2009). "Non-LTR retrotransposons encode noncanonical RRM domains in their first open reading frame." Proceedings of the National Academy of Sciences **106**(3): 731-736.

Kopera, H. C., J. B. Moldovan, T. A. Morrish, J. L. Garcia-Perez and J. V. Moran (2011). "Similarities between long interspersed element-1 (LINE-1) reverse transcriptase and telomerase." Proc Natl Acad Sci U S A, 2011 **108**(51): 20345-20350.

Kubo, S., C. Seleme Mdel, H. S. Soifer, J. L. Perez, J. V. Moran, H. H. Kazazian, Jr. and N. Kasahara (2006). "L1 retrotransposition in nondividing and primary human somatic cells." Proc Natl Acad Sci U S A **103**(21): 8036-8041.

Kulpa, D. A. and J. V. Moran (2005). "Ribonucleoprotein particle formation is necessary but not sufficient for LINE-1 retrotransposition." Hum Mol Genet **14**(21): 3237-3248.

Kulpa, D. A. and J. V. Moran (2006). "Cis-preferential LINE-1 reverse transcriptase activity in ribonucleoprotein particles." Nat Struct Mol Biol **13**(7): 655-660.

Kuroda, M. I., M. J. Kernan, R. Kreber, B. Ganetzky and B. S. Baker (1991). "The maleless protein associates with the X chromosome to regulate dosage compensation in Drosophila." Cell **66**(5): 935-947.

Lander, E. S., L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh, R. Funke, D. Gage, K. Harris, A. Heaford, J. Howland, L. Kann, J. Lehoczky, R. LeVine, P. McEwan, K. McKernan, J. Meldrim, J. P. Mesirov, C. Miranda, W. Morris, J. Naylor, C. Raymond, M. Rosetti, R. Santos, A. Sheridan, C. Sougnez, N. Stange-Thomann, N. Stojanovic, A. Subramanian, D. Wyman, J. Rogers, J. Sulston, R. Ainscough, S. Beck, D. Bentley, J. Burton, C. Clee, N. Carter, A. Coulson, R. Deadman, P. Deloukas, A. Dunham, I. Dunham, R. Durbin, L. French, D. Grafham, S. Gregory, T. Hubbard, S. Humphray, A. Hunt, M. Jones, C. Lloyd, A. McMurray, L. Matthews, S. Mercer, S. Milne, J. C. Mullikin, A. Mungall, R. Plumb, M. Ross, R. Shownkeen, S. Sims, R. H. Waterston, R. K. Wilson, L. W. Hillier, J. D. McPherson, M. A. Marra, E. R. Mardis, L. A. Fulton, A. T. Chinwalla, K. H. Pepin, W. R. Gish, S. L. Chissoe, M. C. Wendl, K. D. Delehaunty, T. L. Miner, A. Delehaunty, J. B. Kramer, L. L. Cook, R. S. Fulton, D. L. Johnson, P. J. Minx, S. W. Clifton, T. Hawkins, E. Branscomb, P. Predki, P. Richardson, S. Wenning, T. Slezak, N. Doggett, J. F. Cheng, A. Olsen, S. Lucas, C. Elkin, E. Uberbacher, M. Frazier, R. A. Gibbs, D. M. Muzny, S. E. Scherer, J. B. Bouck, E. J. Sodergren, K. C. Worley, C. M. Rives, J. H. Gorrell, M. L. Metzker, S. L. Naylor, R. S. Kucherlapati, D. L. Nelson, G. M. Weinstock, Y. Sakaki, A. Fujiyama, M. Hattori, T. Yada, A. Toyoda, T. Itoh, C. Kawagoe, H. Watanabe, Y. Totoki, T. Taylor, J. Weissenbach, R. Heilig, W. Saurin, F. Artiguenave, P. Brottier, T. Bruls, E. Pelletier, C. Robert, P. Wincker, D. R. Smith, L. Doucette-Stamm, M. Rubenfield, K. Weinstock, H. M. Lee, J. Dubois, A. Rosenthal, M. Platzer, G. Nyakatura, S. Taudien, A. Rump, H. Yang, J. Yu, J. Wang, G. Huang, J. Gu, L. Hood, L. Rowen, A. Madan, S. Qin, R. W. Davis, N. A. Federspiel, A. P. Abola, M. J. Proctor, R. M. Myers, J. Schmutz, M. Dickson, J. Grimwood, D. R. Cox, M. V. Olson, R. Kaul, C. Raymond, N. Shimizu, K. Kawasaki, S. Minoshima, G. A. Evans, M. Athanasiou, R. Schultz, B. A. Roe, F. Chen, H. Pan, J. Ramser, H. Lehrach, R. Reinhardt, W. R. McCombie, M. de la Bastide, N. Dedhia, H. Blocker, K. Hornischer, G. Nordsiek, R. Agarwala, L. Aravind, J. A. Bailey, A. Bateman, S. Batzoglou, E. Birney, P. Bork, D. G. Brown, C. B. Burge, L. Cerutti, H. C. Chen, D. Church, M. Clamp, R. R. Copley, T. Doerks, S. R. Eddy, E. E. Eichler, T. S. Furey, J. Galagan, J. G. Gilbert, C. Harmon, Y. Hayashizaki, D. Haussler, H. Hermjakob, K. Hokamp, W. Jang, L. S. Johnson, T. A. Jones, S. Kasif, A. Kaspryzk, S. Kennedy, W. J. Kent, P. Kitts, E. V. Koonin, I. Korf, D. Kulp, D. Lancet, T. M. Lowe, A. McLysaght, T. Mikkelsen, J. V. Moran, N. Mulder, V. J. Pollara, C. P. Ponting, G. Schuler, J. Schultz, G. Slater, A. F. Smit, E. Stupka, J. Szustakowski, D. Thierry-Mieg, J. Thierry-Mieg, L. Wagner, J. Wallis, R. Wheeler, A. Williams, Y. I. Wolf, K. H. Wolfe, S. P. Yang, R. F. Yeh, F. Collins, M. S. Guyer, J. Peterson, A. Felsenfeld, K. A. Wetterstrand, A. Patrinos, M. J. Morgan, P. de Jong, J. J. Catanese, K. Osoegawa, H. Shizuya, S. Choi and Y. J. Chen (2001). "Initial sequencing and analysis of the human genome." Nature **409**(6822): 860-921.

Leeds, P., S. W. Peltz, A. Jacobson and M. R. Culbertson (1991). "The product of the yeast UPF1 gene is required for rapid turnover of mRNAs containing a premature translational termination codon." Genes Dev **5**(12A): 2303-2314.

Leibold, D. M., G. D. Swergold, M. F. Singer, R. E. Thayer, B. A. Dombroski and T. G. Fanning (1990). "Translation of LINE-1 DNA elements in vitro and in human cells." Proc Natl Acad Sci U S A **87**(18): 6990-6994.

Leung, A. K., S. Vyas, J. E. Rood, A. Bhutkar, P. A. Sharp and P. Chang (2011). "Poly(ADP-ribose) regulates stress responses and microRNA activity in the cytoplasm." Mol Cell **42**(4): 489-499.

Levin, H. L. and J. V. Moran (2011). "Dynamic interactions between transposable elements and their hosts." Nature Reviews Genetics **12**(9): 615-627.

Li, X., J. Zhang, R. Jia, V. Cheng, X. Xu, W. Qiao, F. Guo, C. Liang and S. Cen (2013). "The MOV10 helicase inhibits LINE-1 mobility." J Biol Chem **288**(29): 21148-21160.

Licklider, L. J., C. C. Thoreen, J. Peng and S. P. Gygi (2002). "Automation of Nanoscale Microcapillary Liquid Chromatography-Tandem Mass Spectrometry with a Vented Column." Analytical Chemistry **74**(13): 3076-3083.

Liu, J., M. A. Valencia-Sanchez, G. J. Hannon and R. Parker (2005). "MicroRNA-dependent localization of targeted mRNAs to mammalian P-bodies." Nat Cell Biol **7**(7): 719-723.

Liu, L., G. Chen, X. Ji and G. Gao (2004). "ZAP is a CRM1-dependent nucleocytoplasmic shuttling protein." Biochemical and Biophysical Research Communications **321**(3): 517-523.

Luan, D. D., M. H. Korman, J. L. Jakubczak and T. H. Eickbush (1993). "Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition." Cell **72**(4): 595-605.

Mandal, P. K., A. D. Ewing, D. C. Hancks and H. H. Kazazian, Jr. (2013). "Enrichment of processed pseudogene transcripts in L1-ribonucleoprotein particles." Hum Mol Genet **22**(18): 3730-3748.

Mao, R., H. Nie, D. Cai, J. Zhang, H. Liu, R. Yan, A. Cuconati, T. M. Block, J. T. Guo and H. Guo (2013). "Inhibition of hepatitis B virus replication by the host zinc finger antiviral protein." PLoS Pathog **9**(7): e1003494.

Martin, S. L. (1991). "Ribonucleoprotein particles with LINE-1 RNA in mouse embryonal carcinoma cells." Mol Cell Biol **11**(9): 4804-4807.

Martin, S. L. and F. D. Bushman (2001). "Nucleic acid chaperone activity of the ORF1 protein from the mouse LINE-1 retrotransposon." Mol Cell Biol **21**(2): 467-475.

Martin, S. L., M. Cruceanu, D. Branciforte, P. Wai-Lun Li, S. C. Kwok, R. S. Hodges and M. C. Williams (2005). "LINE-1 retrotransposition requires the nucleic acid chaperone activity of the ORF1 protein." J Mol Biol **348**(3): 549-561.

Mathias, S. L., A. F. Scott, H. H. Kazazian, Jr., J. D. Boeke and A. Gabriel (1991). "Reverse transcriptase encoded by a human transposable element." Science **254**(5039): 1808-1810.

McMillan, J. P. and M. F. Singer (1993). "Translation of the human LINE-1 element, L1Hs." Proc Natl Acad Sci U S A **90**(24): 11533-11537.

Miki, Y., I. Nishisho, A. Horii, Y. Miyoshi, J. Utsunomiya, K. W. Kinzler, B. Vogelstein and Y. Nakamura (1992). "Disruption of the APC Gene by a Retrotransposal Insertion of L1 Sequence in a Colon Cancer." Cancer Research **52**(3): 643-645.

Moldovan, J. B. and J. V. Moran, "The Zinc-Finger Antiviral Protein ZAP Inhibits LINE and Alu Retrotransposition." PLoS Genet, 2015 **11**(5): e1005121.

Moran, J. V., S. E. Holmes, T. P. Naas, R. J. DeBerardinis, J. D. Boeke and H. H. Kazazian, Jr. (1996). "High frequency retrotransposition in cultured mammalian cells." Cell **87**(5): 917-927.

Morrish, T. A., N. Gilbert, J. S. Myers, B. J. Vincent, T. D. Stamato, G. E. Taccioli, M. A. Batzer and J. V. Moran (2002). "DNA repair mediated by endonuclease-independent LINE-1 retrotransposition." Nat Genet **31**(2): 159-165.

Muckenfuss, H., M. Hamdorf, U. Held, M. Perkovic, J. Lower, K. Cichutek, E. Flory, G. G. Schumann and C. Munk (2006). "APOBEC3 proteins inhibit human LINE-1 retrotransposition." J Biol Chem **281**(31): 22161-22172.

Muller, S., P. Moller, M. J. Bick, S. Wurr, S. Becker, S. Gunther and B. M. Kummerer (2007). "Inhibition of filovirus replication by the zinc finger antiviral protein." J Virol **81**(5): 2391-2400.

Nesvizhskii, A. I., A. Keller, E. Kolker and R. Aebersold (2003). "A Statistical Model for Identifying Proteins by Tandem Mass Spectrometry." Analytical Chemistry **75**(17): 4646-4658.

Ostertag, E. M., J. L. Goodier, Y. Zhang and H. H. Kazazian, Jr. (2003). "SVA elements are nonautonomous retrotransposons that cause disease in humans." Am J Hum Genet **73**(6): 1444-1451.

Ostertag, E. M., E. T. Prak, R. J. DeBerardinis, J. V. Moran and H. H. Kazazian, Jr. (2000). "Determination of L1 retrotransposition kinetics in cultured cells." Nucleic Acids Res **28**(6): 1418-1423.

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

Perepelitsa-Belancio, V. and P. Deininger (2003). "RNA truncation by premature polyadenylation attenuates human mobile element activity." Nat Genet **35**(4): 363-366.

Raiz, J., A. Damert, S. Chira, U. Held, S. Klawitter, M. Hamdorf, J. Lower, W. H. Stratling, R. Lower and G. G. Schumann (2012). "The non-autonomous retrotransposon SVA is trans-mobilized by the human LINE-1 protein machinery." Nucleic Acids Res **40**(4): 1666-1683.

Rauch, A., M. Bellew, J. Eng, M. Fitzgibbon, T. Holzman, P. Hussey, M. Igra, B. Maclean, C. W. Lin, A. Detter, R. Fang, V. Faca, P. Gafken, H. Zhang, J. Whitaker, D. States, S. Hanash, A. Paulovich and M. W. McIntosh (2005). "Computational Proteomics Analysis System (CPAS): An Extensible, Open-Source Analytic System for Evaluating and Publishing Proteomic Data and High Throughput Biological Experiments." Journal of Proteome Research **5**(1): 112-121.

Reineke, L. C. and R. E. Lloyd (2013). "Diversion of stress granules and P-bodies during viral infection." Virology **436**(2): 255-267.

Richardson, S. R., I. Narvaiza, R. A. Planegger, M. D. Weitzman and J. V. Moran (2014). "APOBEC3A deaminates transiently exposed single-strand DNA during LINE-1 retrotransposition." Elife **3**: e02008.

Sassaman, D. M., B. A. Dombroski, J. V. Moran, M. L. Kimberland, T. P. Naas, R. J. DeBerardinis, A. Gabriel, G. D. Swergold and H. H. Kazazian, Jr. (1997). "Many human L1 elements are capable of retrotransposition." Nat Genet **16**(1): 37-43.

Sawyer, S. L., M. Emerman and H. S. Malik (2004). "Ancient adaptive evolution of the primate antiviral DNA-editing enzyme APOBEC3G." PLoS Biol **2**(9): E275.

Sawyer, S. L. and H. S. Malik (2006). "Positive selection of yeast nonhomologous end-joining genes and a retrotransposon conflict hypothesis." Proc Natl Acad Sci U S A **103**(47): 17614-17619.

Sawyer, S. L., L. I. Wu, M. Emerman and H. S. Malik (2005). "Positive selection of primate TRIM5alpha identifies a critical species-specific retroviral restriction domain." Proc Natl Acad Sci U S A **102**(8): 2832-2837.

Schneider, C. A., W. S. Rasband and K. W. Eliceiri (2012). "NIH Image to ImageJ: 25 years of image analysis." Nature Methods **9**(7): 671-675.

Schumann, G. G. (2007). "APOBEC3 proteins: major players in intracellular defence against LINE-1-mediated retrotransposition." Biochem Soc Trans **35**(Pt 3): 637-642.

Scott, A. F., B. J. Schmeckpeper, M. Abdelrazik, C. T. Comey, B. O'Hara, J. P. Rossiter, T. Cooley, P. Heath, K. D. Smith and L. Margolet (1987). "Origin of the human L1 elements: proposed progenitor genes deduced from a consensus DNA sequence." Genomics **1**(2): 113-125.

Shevchenko, A., M. Wilm, O. Vorm and M. Mann (1996). "Mass Spectrometric Sequencing of Proteins from Silver-Stained Polyacrylamide Gels." Analytical Chemistry **68**(5): 850-858.

Shi, X., A. Seluanov and V. Gorbunova (2007). "Cell divisions are required for L1 retrotransposition." Mol Cell Biol **27**(4): 1264-1270.

Shukla, R., K. R. Upton, M. Munoz-Lopez, D. J. Gerhardt, M. E. Fisher, T. Nguyen, P. M. Brennan, J. K. Baillie, A. Collino, S. Ghisletti, S. Sinha, F. Iannelli, E. Radaelli, A. Dos Santos, D. Rapoud, C. Guettier, D. Samuel, G. Natoli, P.

Carninci, F. D. Ciccarelli, J. L. Garcia-Perez, J. Faivre and G. J. Faulkner (2013). "Endogenous retrotransposition activates oncogenic pathways in hepatocellular carcinoma." Cell **153**(1): 101-111.

Siomi, M. C., K. Sato, D. Pezic and A. A. Aravin (2011). "PIWI-interacting small RNAs: the vanguard of genome defence." Nat Rev Mol Cell Biol **12**(4): 246-258.

Smit, A. F., G. Toth, A. D. Riggs and J. Jurka (1995). "Ancestral, mammalian-wide subfamilies of LINE-1 repetitive sequences." J Mol Biol **246**(3): 401-417.

Stetson, D. B., J. S. Ko, T. Heidmann and R. Medzhitov (2008). "Trex1 prevents cell-intrinsic initiation of autoimmunity." Cell **134**(4): 587-598.

Sugano, T., M. Kajikawa and N. Okada (2006). "Isolation and characterization of retrotransposition-competent LINEs from zebrafish." Gene **365**(0): 74-82.

Suzuki, J., K. Yamaguchi, M. Kajikawa, K. Ichiyanagi, N. Adachi, H. Koyama, S. Takeda and N. Okada (2009). "Genetic evidence that the non-homologous end-joining repair pathway is involved in LINE retrotransposition." PLoS Genet **5**(4): e1000461.

Swergold, G. D. (1990). "Identification, characterization, and cell specificity of a human LINE-1 promoter." Mol Cell Biol **10**(12): 6718-6729.

Taylor, M. S., J. Lacava, P. Mita, K. R. Molloy, C. R. Huang, D. Li, E. M. Adney, H. Jiang, K. H. Burns, B. T. Chait, M. P. Rout, J. D. Boeke and L. Dai (2013). "Affinity Proteomics Reveals Human Host Factors Implicated in Discrete Stages of LINE-1 Retrotransposition." Cell **155**(5): 1034-1048.

Tourriere, H., K. Chebli, L. Zekri, B. Courselaud, J. M. Blanchard, E. Bertrand and J. Tazi (2003). "The RasGAP-associated endoribonuclease G3BP assembles stress granules." J Cell Biol **160**(6): 823-831.

Ullu, E. and C. Tschudi (1984). "Alu sequences are processed 7SL RNA genes." Nature **312**(5990): 171-172.

Wallace, M. R., L. B. Andersen, A. M. Saulino, P. E. Gregory, T. W. Glover and F. S. Collins (1991). "A de novo Alu insertion results in neurofibromatosis type 1." Nature **353**(6347): 864-866.

Weber, M. J. (2006). "Mammalian Small Nucleolar RNAs Are Mobile Genetic Elements." PLoS Genet **2**(12): e205.

Wei, W., N. Gilbert, S. L. Ooi, J. F. Lawler, E. M. Ostertag, H. H. Kazazian, J. D. Boeke and J. V. Moran (2001). "Human L1 retrotransposition: cis preference versus trans complementation." Mol Cell Biol **21**(4): 1429-1439.

Wei, W., T. A. Morrish, R. S. Alisch and J. V. Moran (2000). "A transient assay reveals that cultured human cells can accommodate multiple LINE-1 retrotransposition events." Anal Biochem **284**(2): 435-438.

Wissing, S., M. Montano, J. L. Garcia-Perez, J. V. Moran and W. C. Greene (2011). "Endogenous APOBEC3B Restricts LINE-1 Retrotransposition in

Transformed Cells and Human Embryonic Stem Cells." Journal of Biological Chemistry **286**(42): 36427-36437.

Xie, Y., L. Mates, Z. Ivics, Z. Izsvak, S. L. Martin and W. An (2013). "Cell division promotes efficient retrotransposition in a stable L1 reporter cell line." Mob DNA **4**(1): 10.

Yang, L., J. Brunsfeld, L. Scott and H. Wichman (2014). "Reviving the dead: history and reactivation of an extinct l1." PLoS Genet **10**(6): e1004395.

Yoder, J. A., C. P. Walsh and T. H. Bestor (1997). "Cytosine methylation and the ecology of intragenomic parasites." Trends Genet **13**(8): 335-340.

Zeuthen, J., J. O. Norgaard, P. Avner, M. Fellous, J. Wartiovaara, A. Vaheri, A. Rosen and B. C. Giovanella (1980). "Characterization of a human ovarian teratocarcinoma-derived cell line." Int J Cancer **25**(1): 19-32.

Zhang, A., B. Dong, A. J. Doucet, J. B. Moldovan, J. V. Moran and R. H. Silverman (2014). "RNase L restricts the mobility of engineered retrotransposons in cultured human cells." Nucleic Acids Res **42**(6): 3803-3820.

Zhao, K., J. Du, X. Han, J. L. Goodier, P. Li, X. Zhou, W. Wei, S. L. Evans, L. Li, W. Zhang, L. E. Cheung, G. Wang, H. H. Kazazian, Jr. and X. F. Yu (2013). "Modulation of LINE-1 and Alu/SVA retrotransposition by Aicardi-Goutieres syndrome-related SAMHD1." Cell Rep **4**(6): 1108-1115.

Zhu, Y., G. Chen, F. Lv, X. Wang, X. Ji, Y. Xu, J. Sun, L. Wu, Y. T. Zheng and G. Gao (2011). "Zinc-finger antiviral protein inhibits HIV-1 infection by selectively targeting multiply spliced viral mRNAs for degradation." Proc Natl Acad Sci U S A **108**(38): 15834-15839.

Zhu, Y., X. Wang, S. P. Goff and G. Gao (2012). "Translational repression precedes and is required for ZAP-mediated mRNA decay." EMBO J **31**(21): 4236-4246.