The *trans* effect of non-functional L1 loci on retrotransposition of functional L1 elements

An Abstract

Submitted this 21st day of July, 2017

To the Graduate program in Biomedical Sciences

In Partial Fulfillment of the Requirements

of the School of Medicine

of Tulane University

for the Degree of

Doctor of Philosophy

by

Mark Sokolowski, Jr.

Approved by:

Victoria P. Belancio, PhD (Chair)

David E. Blask, PhD, MD

Prescott L. Deininger, PhD

Erik K. Flemington, PhD

Steven M. Hill, PhD

Abstract

Long Interspersed Element 1 (LINE-1, L1) is an autonomous non-LTR retroelement that is currently active in human and mammalian genomes. A full-length LINE-1 contains a 5' UTR, two open reading frames (ORF1 and ORF2) and a 3' UTR. Translation of L1 mRNA generates two proteins, ORF1 and ORF2, both of which are required for L1 retrotransposition. ORF1 and ORF2 proteins associate with their parental mRNA through a process known as *cis* preference to form RNA/protein complexes required for retrotransposition. L1 has accumulated to about 500,000 copies distributed throughout the genome, of which several thousands are full-length and about 100 are functional, capable of retrotransposition. Consequently, functional and non-functional L1 elements co-exist in mammalian genomes. Many of the full-length L1 loci have accumulated stop codons within their ORF1 sequence. These loci have the potential to produce truncated ORF1p with a yet unknown impact on retrotransposition.

Here we demonstrate that full-length ORF1p monomers generated from different expression plasmids can form heterocomplexes in mammalian cells. This *trans* association suggests that truncated ORF1 proteins may also interact *in trans*. By testing the *trans* effect of the full-length and truncated ORF1p on L1 retrotransposition in mammalian cells, we found that human and mouse ORF1 proteins have differential effects on retrotransposition of their respective L1 elements in human and mouse cells. This effect requires an N-terminus and coiled-coil domain of ORF1p. We demonstrate that a genetic disruption of the leucine zipper motif in the coil-coil domain of human ORF1p abolishes the suppressive effect on L1 retrotransposition. We also generated L1

elements containing mutations resulting in stop codons at specific positions identified by bioinformatic analysis of L1 loci residing in the human and mouse genomes. We found that some of the L1 loci containing stop codons express truncated ORF1 proteins and suppress retrotranspsoition of a functional L1 element. Taken together these findings suggest that L1 retrotransposition may be influenced by coexpression of defective L1 loci and that these L1 loci may reduce accumulation of *de novo* L1 integration events.

The *trans* effect of non-functional L1 loci on retrotransposition of functional L1 elements

An Abstract

Submitted this 21st day of July, 2017

To the Graduate program in Biomedical Sciences

In Partial Fulfillment of the Requirements

of the School of Medicine

of Tulane University

for the Degree of

Doctor of Philosophy

by

Mark Sokolowski, Jr.

Approved by:

Victoria P. Belancio, PhD (Chair)

David E. Blask, PhD, MD

Prescott L. Deininger, PhD

Erik K. Flemington, PhD

Steven M. Hill, PhD

© Copyright by Mark Sokolowski, Jr., 2017

All Rights Reserved

Acknowledgments

I would like to thank my mentor Dr. Victoria P. Belancio for her support and patience, which allowed me the freedom to pursue topics of interest to me. This freedom has greatly increased my independence and skills as a scientist. I would also like to thank members of the Consortium of Mobile Elements at Tulane for their friendship and support. In particular, I would like to thank Dr. Kristine J Kines, Dr. Maria E. Morales, Dr. Cecily B. Defreece, Dawn deHaro, Dr. Geraldine Servant, Dr. Vincent A. Streva, Dr. Clay M. Christian, Ben Freeman and Tiffany Kaul for their discussions, ideas, help and guidance over the years. Also, I thank my committee Dr. David E. Blask, Dr. Prescott L. Deininger, Dr. Erik K. Flemington and Dr. Steven M. Hill for pushing me to excel. Without them, this dissertation would have been a more difficult journey.

I would like to thank my mom, dad and my brother for believing in me, and for building a foundation in education that has allowed me to pursue a doctorate. Without their love, patience and belief in me, I would not have come this far.

TABLE OF CONTENTS

Acknowledgments ii

List of Tables and Figures iii

Chapter 1: Introduction 1

1.1: Long Interspersed Element-1 composition and replication cycle in the human genome 1

1.1.1: L1 ORF1 Protein 4

Chapter 2: Development of a monoclonal antibody specific to the endonuclease domain of the human LINE-1 ORF2 protein 11

2.1: Abstract 11

2.2: Introduction 12

2.3: Materials and methods 15

2.3.1: Cells 15

2.3.2: Transfections 15

2.3.3: Total protein extraction 16

2.3.4: Western blot analysis 16

2.3.5: Plasmids 18

2.3.6: ORF2p endonuclease purification 18

2.3.7: Monoclonal antibody production 19

2.3.8: The LINE-1 EN cleavage assay 19

2.3.9: The Ape1 EN cleavage assay 20

2.3.10: Annealing Oligonucleotides 21

2.3.11: Alignment of human and mouse ORF2p endonuclease domains 21

2.3.12: Calculation of the number of protein molecules 22

2.4: Results 23

2.4.1: Generation of monoclonal antibody against human L1 ORF2p endonuclease 23

2.4.2: anti-ORF2p monoclonal antibody is specific to the ORF2 protein of human origin 25

2.4.3: anti-ORF2p monoclonal antibody recognizes an epitope which includes amino acid 205 of the human ORF2p endonuclease 27

2.4.4: Sensitivity of the anti-ORF2p monoclonal antibody 30

2.4.5: Monoclonal anti-ORF2p antibody inhibits L1 endonuclease activity in an *in vitro* endonuclease cleavage assay 34

2.5: Discussion 37

2.6: Conclusions 41

Chapter 3: Characterization of L1 ORF1p self-interaction and cellular localization using a mammalian two-hybrid system 42

3.1: Abstract 42

3.2: Introduction 43

3.3: Materials and methods 47

3.3.1: Cells 47

3.3.2: Transfections 47

3.3.3: Protein harvest for western blot analysis 48

3.3.4: Mammalian Two-Hybrid 50

3.3.5: Plasmids 50

3.3.6: Statistical Analysis 51

3.4: Results 52

3.4.1: Mammalian two-hybrid system detects human and mouse ORF1p self-interaction 52

3.4.2: ORF1 protein expression in human and mouse cells 57

3.4.3: Subcellular localization of human and mouse ORF1 proteins 60

3.4.4: Heterodimer formation by the ORF1 proteins expressed from different mRNAs (heterodimerization in trans) 66

3.5: Discussion

Chapter 4: Truncated ORF1 proteins can suppress LINE-1 retrotransposition in *trans* **77**

4.1: Abstract 77

4.2: Introduction 78

4.3: Materials and methods 82

4.3.1: Cells 82

4.3.2: Plasmids 82

4.3.3: Transfections 84

4.3.4: Western blot 85

4.3.4.1: Total protein extraction 85

4.3.4.2: Toxicity Assay 85

4.3.4.3: Nuclear/Cytoplasm fractionation 85

4.3.4.4: Western blot analysis 85

4.3.5: Sucrose cushion of the cytoplasmic extract 86

4.3.6: Co-Immunoprecipitation 86

4.3.7: Bioinformatic Analysis 86

4.4: Results 88

4.4.1: Full-length human ORF1p suppresses human L1 retrotransposition in trans 88

4.4.2: The ORF1p fragment containing the N-terminus and C-C domain is sufficient to suppress human L1

retrotransposition 92

4.4.3: Truncated chimeric ORF1 proteins composed of human and mouse sequences do not suppress human L1 retrotransposition 97

4.4.4: Mutations within the C-C domain of the truncated human ORF1p abolish its ability to suppress human L1 retrotransposition 101

4.4.5: Expression levels and combination of truncated ORF1 proteins influence their trans effect on human L1 retrotransposition 103

4.4.6: Bioinformatic analysis of genomic L1 loci identifies species-specific distribution of stop codon positions 107

4.4.7: L1 loci containing one or more stop codons within the ORF1 sequence are expressed in human cell lines 111

4.4.8: ORF1p generated from L1PA2 and L1PA3 suppress L1 retrotransposition in *trans* 112

4.4.9: Full-length wt human L1 elements containing stop codons in their ORF1 sequence suppress human L1 retrotransposition in trans 115

4.4.10: wt L1 retrotransposition is suppressed in HeLa cells constitutively expressing full-length wt L1 elements containing premature stop codons in their ORF1 sequence 116

4.5: Discussion 117

Chapter 5: Discussion **125**

Appendices

Appendix 1: Information Supplemental to "Chapter 2: Development of a monoclonal antibody specific to the endonuclease domain of the human LINE-1 ORF2 protein" 132

Appendix 2: Information Supplemental to "Chapter 3: Characterization of L1 ORF1p self-interaction and cellular localization using a mammalian two-hybrid system" 137

Appendix 3: Information Supplemental to "Chapter 4: Truncated ORF1 proteins can suppress LINE-1 retrotransposition in *trans*" 147

References **200**

LIST OF FIGURES

Chapter 1: Introduction

Figure 1: Schematic of the human L1 3

Figure 2: Schematic of the human ORF1p 5

Figure 3: Alignment of human and mouse ORF1p 7-8

Chapter 2: Development of a monoclonal antibody specific to the endonuclease domain of the human LINE-1 ORF2 protein

Figure 1: Analysis of bacterially purified human endonuclease 24

Figure 2: Analysis of specificity of the custom anti-human ORF2p monoclonal antibody in human cells 26

Figure 3: Analysis of expression of functional and non-functional human ORF2 protein in human cells 29

Figure 4: Analysis of expression of functional and non-functional full-length and truncated human ORF2 proteins in human cells 32

Figure 5: Analysis of ORF2p generated from a functional wild-type and a functional codon-optimized full-length L1 expression plasmids in 293 cells 33

Figure 6: Analysis of sensitivity of the custom anti-human ORF2p monoclonal antibody 35

Figure 7: Monoclonal anti-ORF2p antibody inhibits L1 endonuclease activity in an in vitro endonuclease cleavage assay 36

Chapter 3: Characterization of L1 ORF1p self-interaction and cellular localization using a mammalian two-hybrid system

Figure 1: Analysis of ORF1p interaction using the M2H system 53

Figure 2: Analysis of M2H ORF1 fusion protein expression in

human and mouse cells 56

Figure 3: Analysis of subcellular localization of ORF1 protein fusions in human and mouse cells 59

Figure 4: Analysis of subcellular localization of hORF1p expressed from vectors containing wild type sequences 62

Figure 5: Analysis of ORF1p interactions in trans 65

Chapter 4: Truncated ORF1 proteins can suppress LINE-1 retrotransposition in *trans*

Figure 1: Full-length human ORF1p suppresses human L1 retrotransposition in *trans* 89

Figure 2: A full-length human ORF1p suppresses retrotransposition of chimeric L1s containing human ORF1 sequence 91

Figure 3: Analysis of truncated ORF1 proteins expression and their potential to suppress L1 retrotransposition 94-95

Figure 4: Mutations in the coiled-coil domain of truncated ORF1 protein abolishes its suppressive effect on L1 retrotransposition 99-100

Figure 5: Expression and effect of truncated ORF1 proteins generated from plasmids containing wild-type L1 sequence on L1 retrotransposition 104

Figure 6: Bioinformatic analysis of stop codon positions in retrotranspositionally-incompetent human L1 loci 106

Figure 7: Human L1 retrotransposition is suppressed

by full-length human L1s containing ORF1p stop codons 109

Figure 8: L1 retrotransposition is suppressed in HeLa cells constitutively expressing a full-length human L1 or full-length human L1s containing stop codons in their ORF1 sequence 114

CHAPTER 1: INTRODUCTION

1.1: LONG INTERSPERSED ELEMENT-1 COMPOSITION AND REPLICATION CYCLE IN THE HUMAN GENOME

Long INterspersed Element-1 (LINE-1, L1) is the only active autonomous non-LTR retrotransposon in the human genome (*10, 11*). L1 amplifies through a copy-and-paste mechanism utilizing an RNA intermediate. This has resulted in the accumulation of over 500,000 L1 copies in the human genome representing about 17% of the total content (*10*). In addition, L1 mobilizes non-autonomous retroelements such as SVA and Alu (*13-15*) as well as pseudogenes (*18*). Overall, L1 is responsible for at least one-third of the total DNA content in human genomes (*10, 11*).

A typical full-length L1 is approximately 6 kilobases long. L1 sequence contains a 5' untranslated region (UTR), ORF1, a short intergenic region, ORF2 and a 3' UTR ending in a polyA site and an associated polyA tail (Figure 1) (*19, 20*). L1 is unique in that it generates a bicistronic mRNA, which encodes for two different proteins, Open Reading Frame 1 (ORF1) and Open Reading Frame 2 (ORF2), (*21*) with substantially differing functional and enzymatic activities. There are about 5,000 full-length L1 elements in the genome, the remaining 495,000 L1 copies are 5' truncated (*10, 22, 23*). The majority of the 5,000 full-length L1 elements are assumed to be inert due to inactivating mutations within the ORF1 and/or ORF2 sequence with only 80-100 L1 loci considered to be active (*22, 24-27*). The L1 replication cycle begins when a full-length L1 mRNA is transcribed from a functional L1 locus. The L1 mRNA is exported to the

cytoplasm where the ORF1p and the ORF2p are translated, and bind to the parental L1 mRNA, a process termed *cis*-preference, forming a ribonucleoprotein (RNP) particle (*5, 28-30*). The ORF1p and the ORF2p are both required for L1 retrotransposition (*21, 28*). The L1 RNP gains access to the nucleus through an unknown mechanism, where the ORF2p endonuclease recognizes and cleaves its target sequence 5'-AATTTT-3' (*31*). The cleaved DNA primes reverse transcription of the L1 mRNA by the reverse transcriptase domain also present in the ORF2p (*32, 33*). This process is known as target-primed reverse transcription (TPRT) (*34, 35*). The end result of the TPRT process is a *de novo* insertion that is either a 5' truncated L1 or a full-length L1 (*10, 24-27*). In order to study ORF2p in the context of the mammalian cellular environment, we generated monoclonal antibodies against the human ORF2 endonuclease domain (Chapter 2). The process of *trans*-complementation, by which functional L1 proteins would rescue retrotranspositionally-incompetent L1 loci, was tested and found to be highly inefficient suggesting that L1 proteins function in *cis* to produce successful L1 integration events (*36*). However, Alu, SVA and pseudogenes require and utilize L1 proteins in order to successfully retrotranspose (*13-15, 18*), this demonstrates that L1 proteins can also function in *trans*.

Figure 1. Schematic of the human L1. A typical full-length L1 is approximately 6kb. The 5' untranslated region (UTR) has an internal Pol II promoter, a primate-specific open reading frame 0 (ORF0) is found in the antisense orientation. L1 is bicistronic, coding for two different proteins: Open reading frame 1 (ORF1) and Open Reading Frame 2 (ORF2). There is an intergenic region in between the ORF1 and ORF2. The 3' UTR ends in a polyA site (pA).

1.1.1: L1 ORF1 PROTEIN

The ORF1p is comprised of an N-terminal domain (N), a coiled-coil domain (C-C), an RNA recognition motif (RRM) and a C-terminal domain (CTD) (Figure 2) (*3, 37-40*). The three reported properties of the ORF1p, homotrimerization (*3, 37, 39, 41-43*), RNA binding (*3, 5, 16, 17, 37*) and nucleic acid chaperone activity (*12, 16, 17, 37*), have been discovered using predominantly *in vitro* approaches. The ORF1p homotrimer formation is mediated through the coiled-coil domain (*3, 37, 39, 41-43*) (Figure 2 and 3). The ORF1p RNA binding ability is mediated through the coiled-coil domain, the RRM and the CTD (*3, 5, 16, 17, 37*). The RRM and CTD of the human and mouse ORF1p are highly conserved (Figure 3). The following residue(s) or a combination of residues have been reported to be important for ORF1p RNA binding *in vitro*: KKKR133/137/140/141AAAA (*3*), R159A (*4*), RRR206/210/211AAA (*3*), I218Y (*4*), R220A (*4*), R235A (*3*), REKG235-238AAAA (*5*), RR261/262AA (*5*) or YPAKLS282-287AAAALA (*5*) (Figure 3). The nucleic acid chaperone functions of the mouse ORF1p are mediated through the RMM and CTD (*12, 16, 17, 37*) (Figure 2 and 3). These functions are assumed to also be important to the human ORF1p role in retrotransposition because the two proteins have significant homology in these regions (Figure 3). The following residue(s) are important for mouse ORF1p nucleic acid chaperone abilities *in vitro*: D159H (*12*), R284A (*17*), R297A (*12*), RR297/298AA (*16*) or Y318A (*17*) (Figure 3).

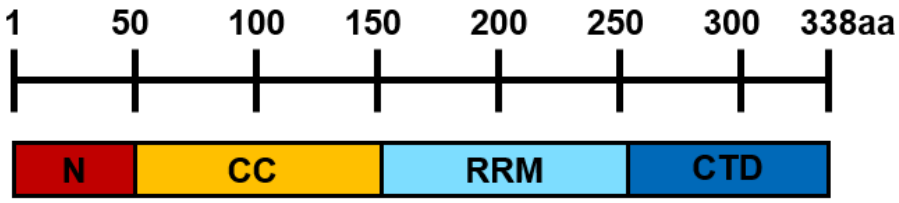

Figure 2. Schematic of the human ORF1p. ORF1 domains are indicated as an N-terminal domain (N), a coiled-coil domain (CC), and RNA recognition motif (RRM) and a C-terminal domain (CTD). Positions of these domains are approximate.

The coiled-coil domain is involved in the trimerization of the ORF1p (*4, 37, 39, 41-43*). It has been demonstrated that three ORF1 monomers come together to form a functional ORF1p homotrimer *in vitro* (*3, 37, 39, 41, 42*). The coiled-coil domain of the human ORF1p is comprised of 14 heptad repeats (*3, 44, 45*), containing both canonical as well as non-conical heptad sequences (*3, 44, 45*). The ORF1p also has three RhxxhE motifs (*3, 45*) located towards the C-terminal end of the coiled-coil domain. These RhxxhE motifs are potentially important for the alignment of parallel trimers (*46*). The N-terminus has been hypothesized to be important to ORF1p trimer-trimer interactions resulting in oligomerization *in vitro* (*37, 42*). Several mutations in the coiled-coil domain (L93/100A) that inhibit successful L1 retrotransposition using engineered L1 elements have been published (*47*), however, the mechanism of this effect has not been elucidated. While studying the L1 proteins *in vitro* allows for precise manipulation of the experimental conditions, it lacks the context of other proteins and pathways that could not only be influencing ORF1p and/or ORF2p function. Recently, over a hundred proteins have been found to interact with ORF1p and/or ORF2p directly or indirectly in human cells (*48, 49*). Although the ORF1 trimerization is assumed to occur in the mammalian cellular environment, its occurrence has never been tested in cultured cells or *in vivo*.

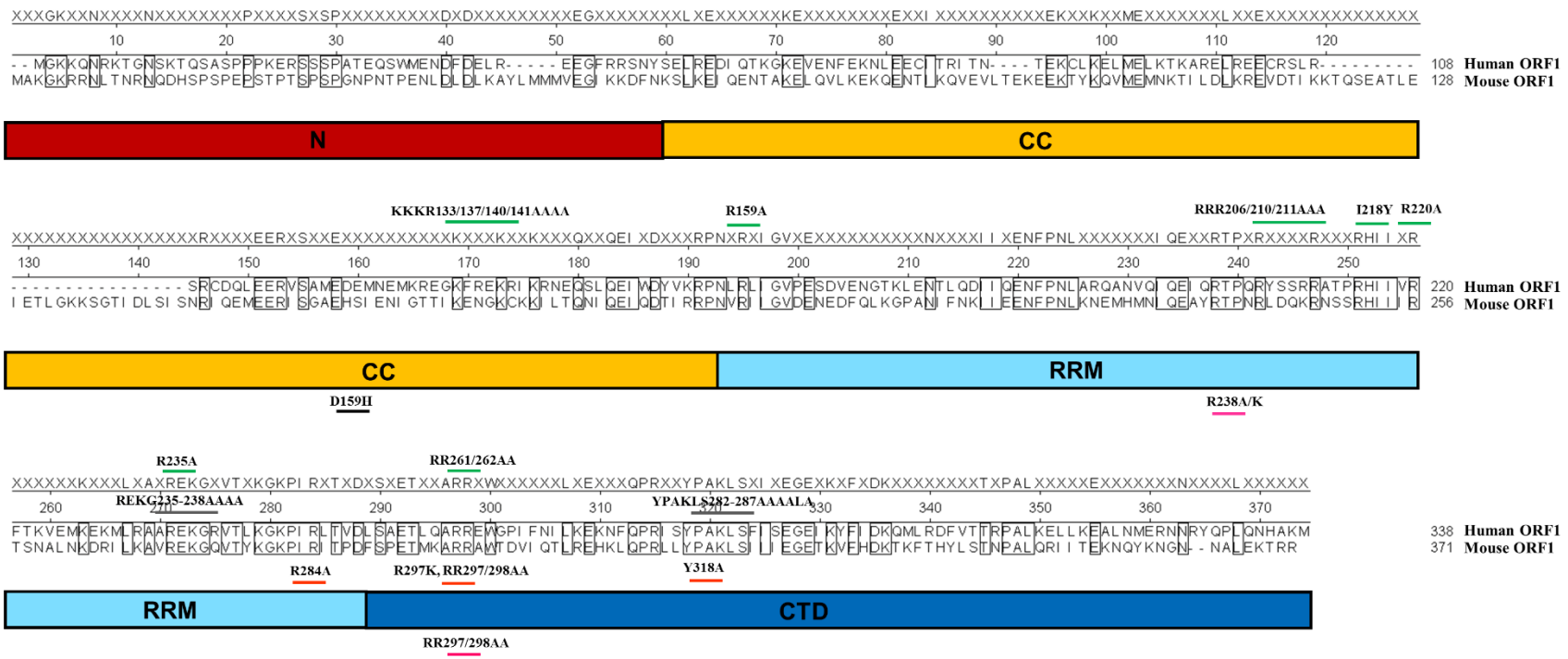
10 20 30 40 50 60 70 80 90 100 110 120
108 Human ORF1
128 Mouse ORF1
N
CC
KKKR133/137/140/141AAAA
R159A
RRR206/210/211AAA
I218Y
R220A
130 140 150 160 170 180 190 200 210 220 230 240 250
220 Human ORF1
256 Mouse ORF1
CC
RRM
D159H
R238A/K
R235A
RR261/262AA
REKG235-238AAAA
YPAKLS282-287AAAALA
260 270 280 290 300 310 320 330 340 350 360 370
338 Human ORF1
371 Mouse ORF1
R284A
R297K, RR297/298AA
Y318A
RRM
CTD
RR297/298AA

Figure 3. Alignment of human and mouse ORF1p. ORF1 domains are indicated as an N-terminal domain (N), a coiled-coil domain (CC), and RNA recognition motif (RRM) and a C-terminal domain (CTD). Positions of these domains are approximate. The Clustal W method in the Lasergene MegAlign 14 program suite was used to align human and mouse ORF1p FASTA amino acid sequence. Residues that are boxed are conserved between human and mouse ORF1p. Green lines denote human ORF1p impaired RNA binding mutants, residues are listed above (*3-5*). The gray lines denote RNA binding mutants of human ORF1 proteins found through L1 RNP/polysome sedimentation, residues are listed above (*5*). Black line denotes a TF_C mouse L1 has a D159, while the TF_{Spa} has a H159, H159 has greatly diminished nucleic acid chaperone abilities compared to D159 (*12*). The orange lines denote mouse ORF1p mutants that had greatly diminished activity in nucleic acid chaperone abilities, residues are listed above (*12, 16, 17*). The pink lines denote RNA binding mutants in mouse ORF1 protein, residues are listed above (*17*).

During the L1 replication cycle, the L1 RNP must gain access to the genomic DNA. There are several potential mechanisms that may be responsible for this necessary step in the L1 replication cycle. The first is an active transport across a nuclear pore to gain access to DNA (*50*), and the second one is a passive access to the genomic DNA during the nuclear envelope breakdown in mitosis (cell cycle) (*51*). Regardless of the mechanism, both ORF1p and ORF2p, tagged or untagged, have been detected in both the nuclear and/or cytoplasm of cells using western blots (*9*) (Chapter 3), immunofluorescence (*47-49, 52, 53*) or immunohistochemistry techniques (*53-58*). Depending on the method and cell type, these results can vary significantly. The discrepancy could be due to the assumption that tagged ORF1p and/or ORF2p behave exactly like their respective untagged counterparts. Our findings demonstrate that the addition of tags to the N-terminus of the human or mouse ORF1p can affect its subcellular localization (Chapter 3).

Both human and mouse genomes harbor hundreds of full-length L1 loci that contain intact ORF1 sequence (*10, 59*) as well as hundreds of L1 loci containing one or more premature stop codons in their ORF1 sequences (*24-27, 60, 61*). If expressed, they can generate either full-length L1 ORF1p or various truncated ORF1 proteins. Even though the impact of these truncated proteins on retrotransposition is unknown there are several reasons to believe that they may negatively impact L1 mobilization because the truncated ORF1 proteins may have a dominant negative effect on L1 retrotransposition. We hypothesize that truncated ORF1 proteins generated by inactive L1 loci inhibit retrotransposition of active L1 loci through *trans* interactions. In order to test this hypothesis in the context of the mammalian cellular environment, we adapted a

mammalian two-hybrid system and non-reducing western blot analysis to determine whether ORF1 proteins generated from different mRNAs can interact with each other in mammalian cells (Chapter 3). Using these approaches, we discovered efficient *trans* interactions between ORF1 proteins translated from different mRNAs. Even though the impact of this *trans*-association on retrotransposition is unknown there are several reasons to believe that this phenomenon may negatively impact L1 mobilization. Our data demonstrate that overexpression of the full-length ORF1p as well as truncated ORF1 proteins suppress L1 mobilization (Chapter 4). Determining the outcome of the coexistence of functional and defective L1 loci in the same genome on L1 retrotransposition is an important step in understanding the role many defective L1 loci may play in maintaining genome stability.

CHAPTER 2: DEVELOPMENT OF A MONOCLONAL ANTIBODY SPECIFIC TO THE ENDONUCLEASE DOMAIN OF THE HUMAN LINE-1 ORF2 PROTEIN

Published as: Mark Sokolowski, Cecily B. DeFreece, Geraldine Servant, Kristine J. Kines, Dawn L. deHaro and Victoria P. Belancio. Development of a monoclonal antibody specific to the endonuclease domain of the human LINE-1 ORF2 protein. *Mobile DNA*. 2014;5(29). doi: 10.1186/s13100-014-0029-x.

2.1: ABSTRACT

Background

LINE-1 (L1) retrotransposons are common occupants of mammalian genomes representing about a fifth of the genetic content. Ongoing L1 retrotransposition in the germ line and somatic tissues has contributed to structural genomic variations and disease-causing mutations in the human genome. L1 mobilization relies on the function of two, self-encoded proteins, ORF1 and ORF2. The ORF2 protein contains two characterized domains: endonuclease and reverse transcriptase.

Results

Using a bacterially purified endonuclease domain of the human L1 ORF2 protein, we have generated a monoclonal antibody specific to the human ORF2 protein. We

determined that the epitope recognized by this monoclonal antibody includes amino acid 205, which is required for the function of the L1 ORF2 protein endonuclease. Using an *in vitro* L1 cleavage assay, we demonstrate that the monoclonal anti-ORF2 protein antibody partially inhibits L1 endonuclease activity without having any effect on the *in vitro* activity of the human AP endonuclease (APE1).

Conclusions

Overall, our data demonstrate that this anti-ORF2 protein monoclonal antibody is a useful tool for human L1-related studies and that it provides a rationale for the development of antibody-based inhibitors of L1-induced damage.

Keywords: LINE-1, ORF2, endonuclease, L1, Retrotransposition, L1 antibody, in vitro assay

2.2: INTRODUCTION

Long interspersed element-1 (L1) is an autonomous non-long terminal repeat retrotransposon that has parasitized the human genome for millions of years. L1 has shaped the evolution of the human genome through a copy-and-paste mobilization of itself (*7*), as well as the short interspersed element (SINE) Alu (*62*), SINE-VNTR-Alu elements (SVA) (*14*), and processed cellular transcripts (*63*). Functional full-length L1 transcripts contain two open reading frames (ORFs) encoding ORF1 and ORF2 proteins (ORF1p and ORF2p, respectively) (Figure 1A). These L1 proteins exhibit *cis*-preference

for their encoding L1 mRNA (*5, 29, 36*) and are utilized in *trans* by the Alu and SVA elements (*14, 15, 62*). L1, Alu, and SVA form ribonucleoprotein (RNP) particles which reach the nucleus to complete their replication cycles by integrating in the host genome via a process of target-primed reverse transcription (*64, 65*). This copy-and-paste process has produced approximately 500,000 L1 loci, accounting for about 17% of the human genome, and over 1,000,000 copies of Alu, which comprise about 11% of our genome (*10*). The majority of the L1 loci are 5' truncated with about 80-100 full-length L1 copies demonstrated to be retrotranspositionally active (*24-27, 66*).

L1 proteins are produced from the full-length L1 mRNA with significantly different efficiencies, mostly owing to the unconventional translation from the bicistronic L1 mRNA (*3, 37, 41, 49*) (Figure 1A). Detection of both L1-encoded proteins is important for understanding L1 biology because they play critical, but different roles in the L1 replication cycle. The human ORF2p is a 149 kilodalton (kDa) protein with three annotated domains: an N-terminal endonuclease (EN) domain (*67*), a reverse transcriptase (RT) domain (*68*), and a C-terminal domain (*69*) with putative RNA binding activity (*70*). Human and mouse L1 ORF2 proteins exhibit a high degree of sequence homology and conservation of function making findings in mouse model systems biologically relevant to the replication cycle of the human L1 (*21, 71*). Although much has been learned about ORF2p function *in vitro* and in mammalian cells using overexpressed tagged ORF2 proteins and polyclonal anti-ORF2p antibodies (*60, 72-74*), having a monoclonal antibody that can detect the untagged human ORF2 protein would be a useful molecular tool to study the requirements for the human L1 ORF2p expression

and activity. It would also aid in advancing our appreciation of the ORF2p impact on host genome stability and in understanding the consequences of its activity to human health.

To satisfy the need for a continuous source of antibodies to detect L1 ORF2p, we developed an anti-ORF2p monoclonal antibody capable of recognizing sequences within the endonuclease domain of the human ORF2 protein. This monoclonal antibody is specific to the human ORF2p and can detect the full-length ORF2 protein, as well as truncated ORF2 proteins overexpressed in mammalian cells. Using a recombinant human L1 endonuclease purified from bacterial cells as a standard (*75*, *76*), we determined the sensitivity of this monoclonal anti-human ORF2p antibody. The unique location of the epitope, encompassing a position required for the function of the human endonuclease domain, allowed us to test the ability of this monoclonal anti-ORF2p antibody to inhibit L1 endonuclease activity *in vitro* using a fluorescence-based cleavage assay.

2.3: MATERIALS AND METHODS

2.3.1: Cells

FLP-In™-293 (Invitrogen) cells were cultured in HyClone Dulbecco's modified Eagle's medium (DMEM) with 10% fetal bovine serum (Invitrogen) and maintained under 6% CO_2 at 37°C. HeLa (ATCC CCL2) cells, NIH-3T3 (ATCC CRL-1658) and Ntera2 (ATCC CRL-1973) cells were maintained as previously described (*77*).

2.3.2: Transfections

Western blot: 293 cells were seeded at 1.5x10^6 cells per T25 flask and transfected 16-18 hours later with 2µg of the human or mouse ORF2 or EN expression plasmids (*60*), or 1, 2 or 4µg of codon-optimized L1Pa1 (*78*) (L1co) or wild-type L1.3 (L1wt) (*36*, *78*). Plus reagent (6µl) (Invitrogen) and Lipofectamine (8µl) were used for each ORF2 or EN transfection reaction in serum-free media. 12µl of plus reagent and 24µl of Lipofectamine were used for each transfection reaction with L1co or L1wt in serum-free media. Transfections with maximum amount of the empty pCDNA plasmid was used as control. After three hours, serum-free media was replaced with serum-containing media, and the cells were harvested at 24 hours after transfection unless otherwise noted in the figure. HeLa and NIH-3T3 cells were seeded at 2x10^6 and 2.5x10^6 cells per T75 flask, respectively, and transfected as previously described using 6µg of plasmid (*9*), 12µl of Plus reagent (Invitrogen) and 18µl and 24 µl of Lipofectamine, respectively, were used in each transfection reaction in serum-free media.

2.3.3: Total protein extraction

Total protein was extracted as previously described (*9*, *60*) using phosphate buffered saline (137mM NaCl (Sigma S9888), 2.7mM KCl (Sigma P4505), 10mM Na_2HPO_4 (Sigma S3264), 2mM KH_2PO_4 (Sigma P9791), pH=7.4), 5mM Ethylenediaminetetraacetic acid (EDTA, Sigma ED), and 0.02% Sodium Azide (Sigma S2002). Lysis buffer was supplemented with phosphatase inhibitors 2 and 3 (Sigma P5726, P0044 respectively) and Halt Protease inhibitors at 10µl/mL each. The samples were subjected to two freeze (-80°C)/thaw (25°C) cycles. The samples were sonicated three times for 10 seconds at 12 watts RMS using a 3mm wide QSonica Microson homogenizer with Microson ultra sonic disruptor XL2000 (Misonix). The protein concentration of each sample was determined using 595nm wavelength OD values against a Bovine Serum Albumin (BSA) standard.

2.3.4: Western blot analysis

10-20µg of total protein were combined with 2x Laemmli buffer, 1.6µl (14.3M) β-mercaptoethanol and boiled for 5 minutes prior to fractionation on Tris Acetate 3-8% Midi gels, Bis Tris 4-12% Midi gels (Invitrogen) and transferred onto nitrocellulose membranes (iBlot System; Invitrogen). Membranes containing fractionated protein samples were blocked for 1 hour in PBS-Tween containing 5% milk and incubated with a 1:250 dilution of custom, polyclonal antibodies against the mouse ORF2p endonuclease (*60*), a 1:500 dilution of custom, polyclonal antibodies against the human ORF2p endonuclease (*60*, *72*) antibodies or a 1:250 dilution of custom monoclonal antibodies

against the human ORF2p endonuclease overnight at 4°C. Detection was carried out using horseradish peroxidase-conjugated secondary antibodies, either HRP-donkey anti-goat (Santa Cruz; sc-2020), HRP-donkey anti-rabbit (Santa Cruz; sc-2317), or HRP-goat anti-mouse (Santa Cruz; sc-2031) at a 1:5000 dilution in 3% milk in PBS-Tween for 1 hour. A 1:5000 dilution of GAPDH antibodies (Santa Cruz sc-25778) was used as an equal loading control. A HRP conjugated monoclonal antibody against the 6x HIS tag (Pierce MA1-21315-HRP) was used at a 1:2000 dilution. All western blots were developed using Clarity™ Western ECL Substrate (Bio-Rad, Cat. #170-5061).

SDS Tris Glycine gels (Figure 5 and Supplemental Figure 4). 3-20µg of total protein were combined with 2x Tris Glycine SDS sample buffer, 1.6µl (14.3M) β-mercaptoethanol and boiled for 5 minutes prior to fractionation on Tris Glycine 4% Mini gels with Tris Glycine SDS running buffer (Invitrogen) and transfer onto nitrocellulose membranes (iBlot System; Invitrogen). Membranes containing fractionated proteins were blocked for 1 hour in PBS-Tween containing 5% milk at room temperature. The membranes were then incubated overnight at 4°C with 1mL of Ab-containing hybridoma supernatant in a blocking mixture containing 4mL of media collected from NIH-3T3 cells cultured for 24 hours and 15mL of 3% milk in PBS-Tween. Detection was carried out using horseradish peroxidase-conjugated secondary antibodies HRP-goat anti-mouse (Santa Cruz; sc-2031) at a 1:5000 dilution in 3% milk in PBS-Tween for 1 hour. All western blots were developed using Clarity™ Western ECL Substrate (Bio-Rad, Cat. #170-5061).

Tris Glycine Native gel (Supplemental Figure 2). 100ng of bacterially purified human ORF2p endonuclease was combined with 2X Native Tris Glycine sample buffer along with 5% GelCode Blue Stain Reagent (Thermo Scientific, Prod # 24592) and fractionated on a Tris Glycine 4-12% gel with Tris Glycine Native running buffer (Invitrogen). Fractionated proteins were transferred onto a nitrocellulose membrane (iBlot System; Invitrogen). Membranes containing fractionated proteins were blocked for 1 hour in PBS-Tween containing 5% milk at room temperature. The membranes were then incubated overnight at $4^{O}C$ with 1mL of Ab-containing hybridoma supernatant in a blocking mixture containing 4mL of media collected from NIH-3T3 cells cultured for 24 hours and 15mL of 3% milk in PBS-Tween. Detection was carried out using horseradish peroxidase-conjugated secondary antibodies HRP-goat anti-mouse (Santa Cruz; sc-2031) at a 1:5000 dilution in 3% milk in PBS-Tween for 1 hour. All western blots were developed using Clarity™ Western ECL Substrate (Bio-Rad, Cat. #170-5061).

2.3.5: Plasmids

All endonuclease constructs used are previously described in (*60*), ORF2 constructs were described in (*60*), L1Pa1 (codon-optimized full-length L1) was described in (*78*). ‘L1wt’ is JM101/L1.3 no tag (*36*).

2.3.6: ORF2p endonuclease purification

A human ORF2 endonuclease was expressed in bacteria and the EN protein was purified as previously described (*75, 76, 79*).

2.3.7: Monoclonal antibody production

hORF2p endonuclease was bacterially purified as previously described (*75, 76*). This purified human ORF2 endonuclease protein was used for immunization of 6 Balb/c mice to generate monoclonal anti-ORF2p antibodies following standard immunization protocol. Briefly, three sequential immunizations (with two week intervals between the injections) with antigen, (purified ORF2p endonuclease diluted in saline) in complete Freund's Adjuvant for the first injection and incomplete Freund's Adjuvant for the second and third injection, injected intrapertionally were performed. The fourth and final immunization was done using the antigen in saline. Mice were bled and tested using ELISA to determine which mouse to use as the source of B-cells for hybridoma production. Electrofusion was preformed between B-cells harvested from the spleen and myeloma cells to produce hybridomas. Resulting hybridoma clones were screened with indirect ELISA to identify positive clones. The final stock of antibody was obtained by protein-G affinity column purification. The antibodies were stored in a PBS with 0.02% W/V sodium azide storage solution. The affinity purified hORF2p monoclonal antibodies were used for subsequent testing.

2.3.8: The LINE-1 EN cleavage assay

The LINE-1 EN was expressed and purified as described previously (*75, 76*).The LINE1 EN cleavage assay was performed using 200 nM purified LINE1 EN, 100 nM of a duplexed oligonucleotide containing LINE-1 EN target site. The reaction buffer

contained 20 mM Hepes (pH 6.5), 150 mM NaCl, 1 mM MgCl2, 1 mM dithiothreotol (DTT), 1% dimethyl sulfoxide (DMSO), 0.1% triton and 0.01 % sodium azide.

The effect of the monoclonal anti-ORF2p antibody on LINE1 EN activity was tested using three concentrations: 100 nM, 150 nM and 200 nM. The antibody was diluted into the above described reaction buffer just prior to use. The same was done for the anti-hORF1p antibody (*9*). A buffer control was used for background subtraction, in which the same volume of buffer alone as the volume of buffer containing antibody was added to the reactions. The LINE-1 EN and APE1 EN cleavage reactions were carried out at 37°C for 30 minutes. The reactions were stopped by quenching on ice and the addition of stop solution: 1X Tris borate EDTA buffer, 80% formamide, 0.01 mM EDTA and xylene cyanol. The samples were run on 18% denaturing acrylamide gels and were analyzed using the Typhoon imager (GE Lifesciences). Fluorescense intensity (FI) was measured using Image Quant software (GE Lifesciences) and graphed using Prism software (GraphPad software, LLC). The percent inhibition of each reaction was determined using the following equation: $\%Inh = 100*(1-(FI_{Antibody}-FI_{Buffer\ Control})/(FI_{L1/APE1\ EN}-FI_{Buffer\ Control}))$.

2.3.9: The Ape1 EN cleavage assay

The purified APE1 EN was purchased from New England Biolabs. The assay was performed using 0.01 and 0.1 units of enzyme and 200 nM of duplexed oligonucleotide containing an abasic site. The sequence of the oligonucleotide was based upon previously published work (*80*). The reaction buffer contained 50mM potassium acetate 20mM tris-

acetate, 10mM magnesium acetate, 1mM DTT, 1% DMSO, 0.1% triton and 0.01% sodium azide.

2.3.10: Annealing oligonucleotides

All oligonucleotides were purchased from Integrated DNA Technologies. The oligonucleotides used in the assays were annealed by adding equivalent amounts of each complimentary nucleotide in annealing buffer (50mM Hepes (pH 7.5) and 100 mM NaCl). The samples were incubated in boiling water for 5 minutes and slow cooled for 1 hour in the dark. The sequence for the LINE1 EN oligonucleotides used in the assay are as follows: 5'/AlexaFluor488/CCTTTTTTTTTAACCGC3' and 5'GCGGTTAAAAAAAAAGG3'. The sequence for the APE1 EN oligonucleotides used in the assay are as follows: 5'/AlexaFluor488/GCCCCC_GGGGACGTACGATATCCCGCTCC3' (where "_" represents an abasic site) and 5'GGAGCGGGATATCGTACGTCCCCCGGGGGC3'.

2.3.11: Alignment of human and mouse ORF2p endonuclease domains

Human L1Pa family consensus sequences (*81*) and L1 Spa (*82*) ORF2 sequence were converted to amino acid sequences and aligned using DNASTAR MegAlign program through the Clustal V method utilizing a gap penalty of '10' and a gap penalty length of '10'.

2.3.12: Calculation of the number of protein molecules

The molecular weight of all proteins was calculated based on their amino acids composition using EditSeq software. The number of molecules detected by monoclonal anti-ORF2p antibody was calculated using the following formula:

X molecules = [**Mass** (g)/**Molecular weight of a specific protein** (g/mol)] X **6.022 x 10^{23}** (mol^{-1})

2.4: RESULTS

2.4.1: Generation of monoclonal antibody against human L1 ORF2p endonuclease

A recombinant human protein containing an ORF2p EN domain N-terminally fused to a His-tag was purified from bacterial cells as previously described (*75*, *76*, *79*), subjected to SDS-PAGE and visualized using coomassie stain (Figure 1B, coomassie panel, expected product of 29 kDa). The efficiency of purification was also confirmed using antibodies against the His-tag fused to the N-terminus of the ORF2p EN (Figure 1B, His-Tag panel). This purified recombinant human EN protein was used for the immunization of Balb/c mice to generate monoclonal anti-ORF2p antibodies following a standard immunization protocol (see material and methods). This approach resulted in a positive hybridoma clone which was used to produce the purified anti-ORF2p monoclonal antibodies. Western blot analysis using this custom ORF2p monoclonal antibody detected a product of the expected size in the clarified lysate and the final elution of the human EN protein used for inoculation (Figure 1B, ORF2 monoclonal panel).

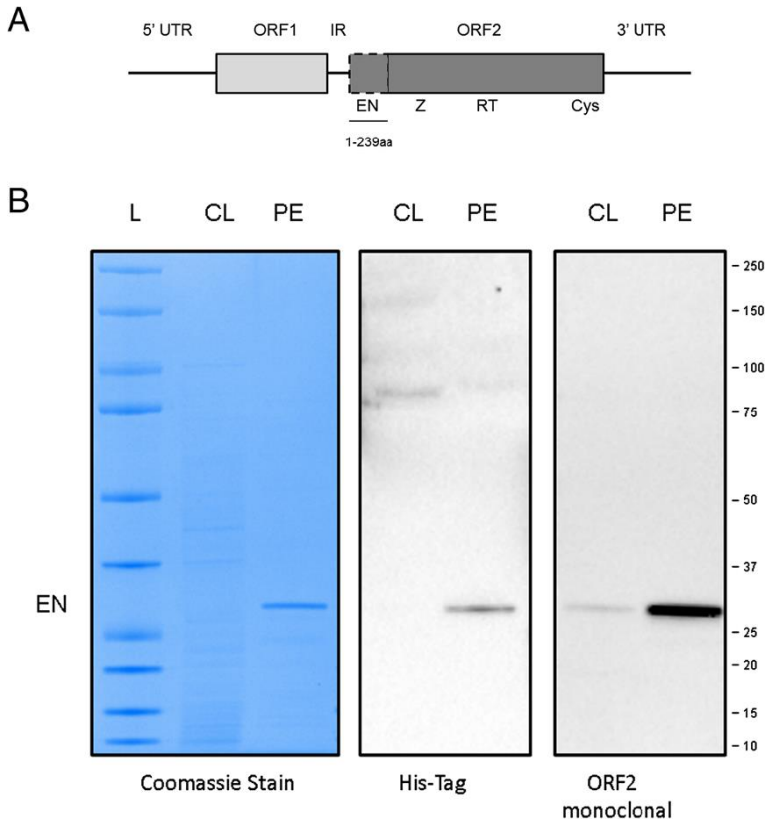

Figure 1. Analysis of bacterially purified human endonuclease. A. Schematic of a full-length L1, which contains a 5' untranslated region (UTR) followed by an ORF1 sequence, an intergenic region, an ORF2 sequence and a 3' UTR. The EN region of the ORF2 sequence subcloned to generate the purified ORF2p endonuclease (EN, 1-239aa) is indicated with a dashed box. **B. (left panel)** Coomassie stain of SDS-PAGE gel. Ladder (L), clarified lysate from bacteria expressing ORF2p endonuclease (CL), and final purified elution (PE) are shown. 500ng of protein was loaded in each lane. **(middle panel)** Western blot analysis of 500ng of CL and PE with HIS-tag specific antibodies. The ORF2p endonuclease used in this study has a HIS-tag fused to its N-terminus (expected size of the His EN protein is 29 kilodaltons, kDa). **(right panel)** Western blot analysis of 500ng of CL and PE with a custom anti-human ORF2p monoclonal antibody. Molecular markers on the right, 10-250 kDa.

2.4.2: anti-ORF2p monoclonal antibody is specific to the ORF2 protein of human origin

We determined that our anti-ORF2p monoclonal antibody detects full-length ORF2p and ORF2p endonuclease in total cell lysates from 293 cells transiently transfected with plasmids containing human codon-optimized full-length ORF2 or the ORF2 endonuclease sequences (Figure 2A, lane hORF2 and hEN). Because the endonuclease domain of the L1 ORF2 protein is highly conserved between the human and mouse ORF2 proteins, we tested whether our antibody discriminates between ORF2 proteins of human and mouse origin. Plasmids encoding mouse codon-optimized full-length ORF2 or ORF2 endonuclease sequences were transiently transfected into 293 cells and total cellular lysates were analyzed by SDS-PAGE followed by immunoblotting with the anti-ORF2p monoclonal antibody. This approach determined that the anti-ORF2p monoclonal antibody does not detected mouse ORF2 or EN proteins (mORF2p and mENp, respectively) even though it detected both human ENp and ORF2p (Figure 2A, monoclonal Ab panel). The mouse ENp and ORF2p were detected when western blot analysis was performed with polyclonal antibodies raised against the endonuclease domain of the mouse ORF2p (*60*) (Figure 2B, mouse Ab panel) confirming that the proteins are expressed under these transfection conditions.

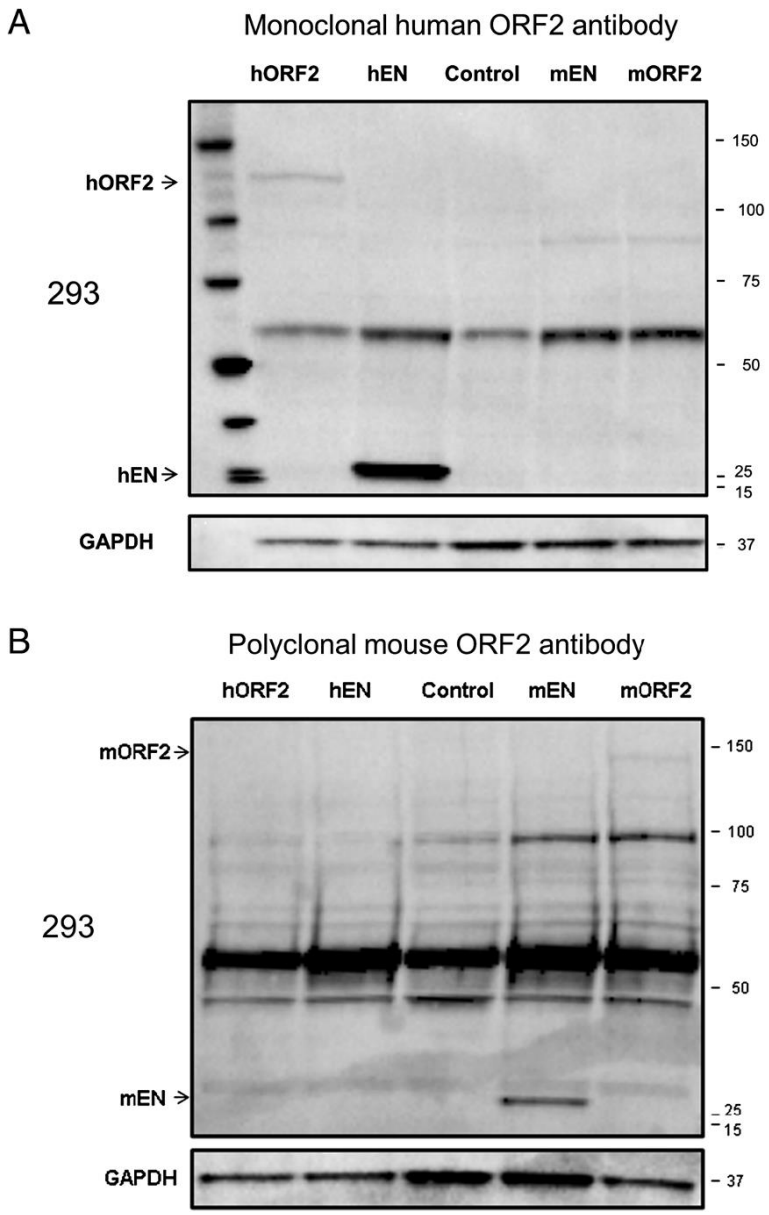

Figure 2. Analysis of specificity of the custom anti-human ORF2p monoclonal antibody in human cells. A. Western blot analysis of mouse and human ORF2 (predicted size 150 and 149 kDa, respectively) and EN (predicted size 30 and 26 kDa, respectively) proteins generated from expression plasmids containing codon-optimized human ORF2 (hORF2), a codon-optimized sequence corresponding to the human ORF2 endonuclease fragment (hEN), codon-optimized mouse ORF2 (mORF2) and a codon-optimized sequences corresponding to the mouse ORF2 endonuclease fragment (mEN) transiently transfected in 293 cells. Custom anti-human ORF2p monoclonal antibody specifically detect proteins of human origin. **B.** Western blot analysis of the same samples as in A was performed with custom anti-mouse ORF2p polyclonal antibodies which specifically detect proteins of mouse origin. Control lane indicates cells transiently transfected with an empty vector. GAPDH is used as a loading control. 15-150 kDa on the right indicate positions of molecular markers.

2.4.3: anti-ORF2p monoclonal antibody recognizes an epitope which includes amino acid 205 of the human ORF2p endonuclease

Many experimental approaches designed to analyze the expression and function of ORF2p involve the use of both functional and non-functional ORF2 proteins. The most commonly used mutations, which abolish the activity of the ORF2p endonuclease are D205A and H230A (*67, 83*). Western blot analysis of total cellular lysates from human and mouse cells transiently transfected with EN or EN 205, 230 plasmids containing codon-optimized sequences producing functional or non-functional (D205A, H230A double mutant) human endonucleases demonstrated that the anti-ORF2p monoclonal antibody detects the active, but not the mutated, endonuclease protein (Figure 3A, monoclonal and Supplementary Figure 1, monoclonal). Both proteins were detected using the polyclonal anti-ORF2p antibody (*60*) (Figure 3A and Supplementary Figure 1).

To determine which EN mutation is responsible for the loss of detection by our anti-ORF2p monoclonal antibody, total cellular lysates from cells transiently transfected with EN 205 and EN 230 plasmids expressing non-functional endonucleases with D205A or H230A mutations were used for western blot analysis with the anti-ORF2p monoclonal antibody (Figure 3B, monoclonal). This approach demonstrated that the anti-ORF2p monoclonal antibody detects ENp containing the H230A mutation, but not the ENp with the D205A mutation (Figure 3B, monoclonal). Both ENp mutants are readily detected with the anti-ORF2p polyclonal antibody (*60, 72*), demonstrating that both proteins are produced under these transfection conditions (Figure 3B, polyclonal). A similar result was obtained when the monoclonal anti-ORF2p antibody was used to detect

transiently expressed functional (ORF2) and non-functional (single and double mutants) full-length human ORF2 proteins (ORF2 205, ORF2 230, and ORF2 205,230, respectively) as well as truncated, functional and double mutant human ORF2 proteins (ENz and ENRT) (*60*) (Figure 4A-E, ORF2, ENz, and ENRT). The anti-ORF2p monoclonal antibody specifically detected functional, but not the non-functional ENz, ENRT, and ORF2 proteins containing the D205A and H230A mutations, even though all of these proteins were produced in these cells as confirmed by western blot analysis using polyclonal anti-ORF2p antibodies (Figure 4D and E). These results support that the epitope recognized by the anti-ORF2p monoclonal antibody includes amino acid 205 of the human ORF2p endonuclease domain.

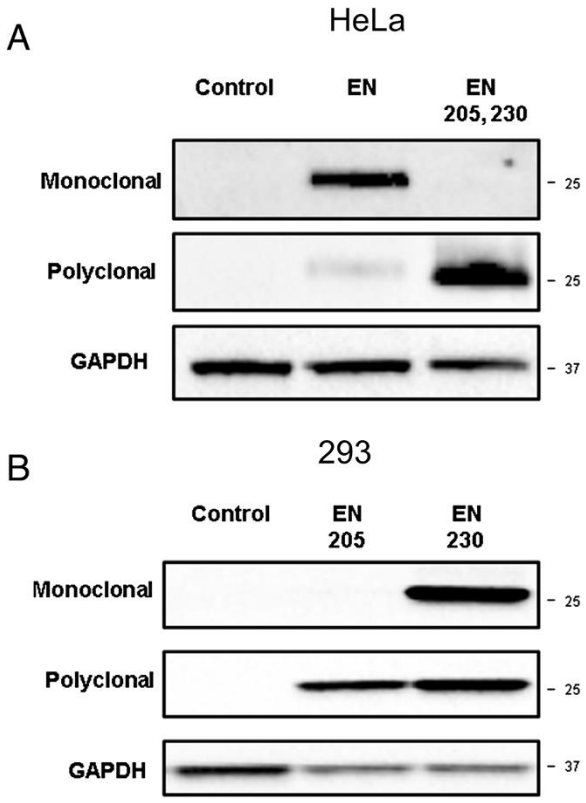

Figure 3. Analysis of expression of functional and non-functional human ORF2 protein in human cells. A. Western blot analysis of proteins generated from expression plasmids containing codon-optimized, functional ORF2 endonuclease sequence (EN) and non-functional ORF2 endonuclease sequence transiently transfected in HeLa cells. Western blot analysis is performed with anti-human ORF2p monoclonal antibody **(top)**, previously described anti-human ORF2p endonuclease polyclonal antibodies **(middle)**, GAPDH antibodies **(bottom)**. The non-functional ORF2 endonuclease sequence (EN 205, 230) has mutations resulting in expression of inactive ENp with D205A and H230A mutations. **B.** The same experiment and analysis as in A, but using 293 cells. Western blot analysis of codon-optimized, non-functional ORF2 endonuclease sequences containing single inactivating mutations D205A or H230A (EN205 and EN 230, respectively) transiently transfected in 293 cells with anti-human ORF2 monoclonal antibody **(top)** or previously described anti-human ORF2p endonuclease polyclonal antibodies **(middle)**. Expected EN protein size is 26 kDa. Control lane indicates cells transiently transfected with an empty vector. 25 and 37 kDa on the right indicate positions of molecular markers.

2.4.4: Sensitivity of the anti-ORF2p monoclonal antibody

The advent of L1 expression plasmids containing codon-optimized sequences greatly facilitated our ability to detect L1-encoded proteins in transfected mammalian cells (*78, 84*). However, it remains important to study L1 proteins generated from wild-type L1 sequences and to understand the difference in expression levels between proteins generated from codon-optimized and wild-type L1 sequences. As with the wild-type full-length L1 and ORF2 expression plasmids, codon-optimized full-length L1 expression plasmids produce much less ORF2 protein than those containing codon-optimized ORF2 sequence (*21, 49*). Consistent with this fact, our anti-ORF2p monoclonal antibody detected different levels of ORF2p in cells transfected under the same conditions with L1 expression plasmids containing wild-type or codon-optimized sequences (*36, 78*) (Figure 5). Transient transfection of 293 cells with plasmids containing codon-optimized ORF2 or full-length wild-type L1 sequences produced the highest and the lowest levels of ORF2 protein, respectively. Transfection of increasing amounts of plasmids containing codon-optimized or wild type full-length L1 sequences demonstrated that detectable levels of ORF2p were observed when 2 and 4µg of the respective plasmids were used. No signal consistent with detection of endogenous ORF2p in 293, Ntera2, nor HeLa cells was observed (Supplementary Figure 2). Using the recombinant human endonuclease purified from bacterial cells as a standard, we determined that 27.6µg of our anti-ORF2p monoclonal antibody is able to detect 10ng ($2.2X10^{17}$ molecules) of the purified hEN under these blotting conditions (Figure 6A). Based on the standard curve generated by western blot analysis of the recombinant human EN purified from bacterial cells, we determined that transfection of expression plasmids containing codon-optimized (co)

human L1 sequences produce 5-6 times more endonuclease protein than that observed from cells transfected with equivalent amounts of plasmids containing wild-type sequences (Figure 6, ENwt and ENco).

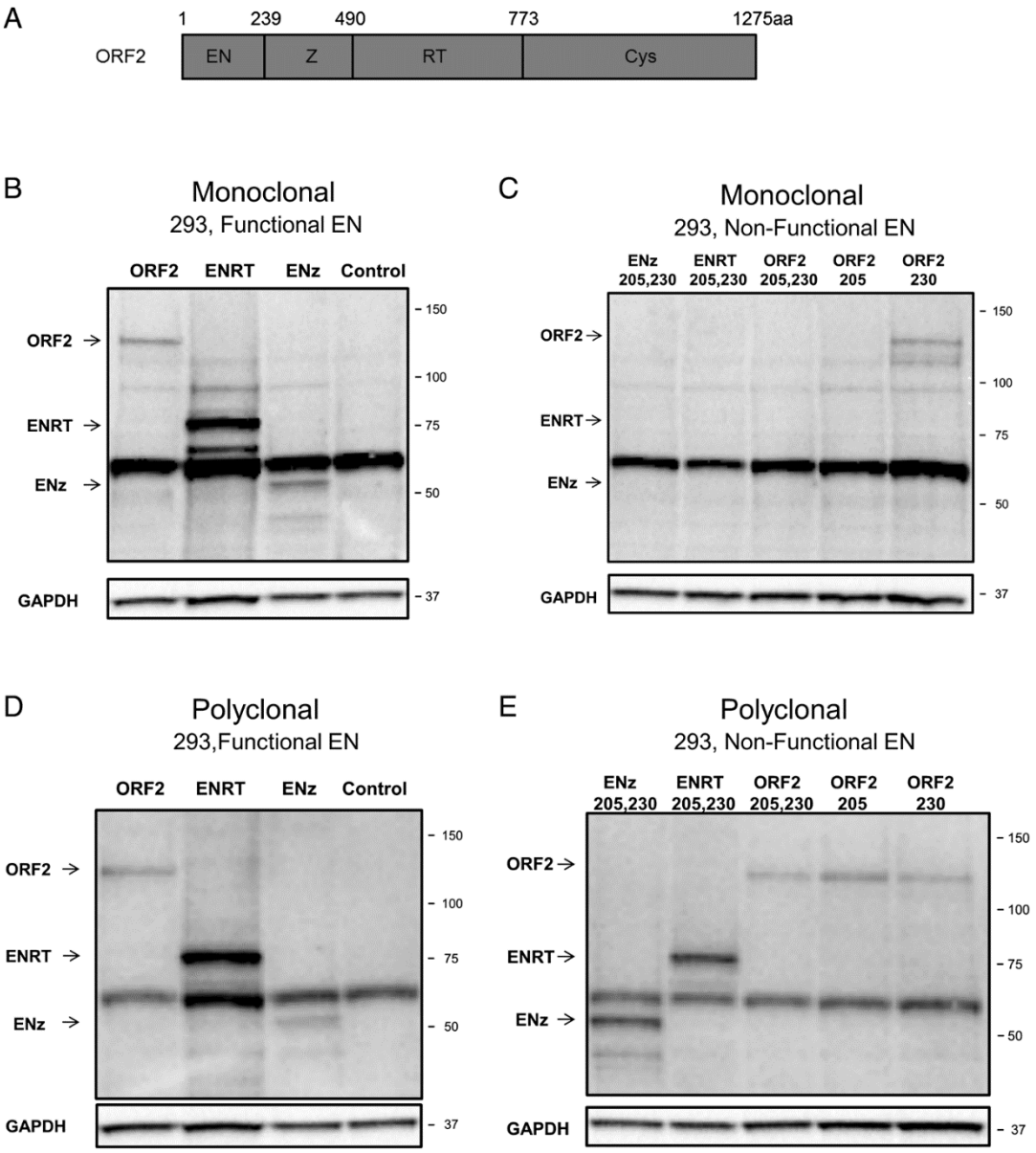

Figure 4. Analysis of expression of functional and non-functional full-length and truncated human ORF2 proteins in human cells. A. Schematic of L1 ORF2 protein. The following ORF2p domains are listed: Endonuclease domain (EN), Z domain (Z), Reverse Transcriptase domain (RT) and the Cysteine-Rich domain (Cys). Amino acid boundaries of each domain are listed. **B.** Western blot analysis of the full-length and truncated ORF2 proteins generated from expression plasmids containing codon-optimized, functional full-length ORF2 (ORF2) and C-terminally truncated ORF2 sequences transiently transfected in 293 cells with monoclonal antibody. Control lane indicates cells transiently transfected with an empty vector. **C.** Western blot analysis of proteins with two inactivating mutations are labeled as 205, 230, single mutants are labeled as 205 or 230 generated from expression plasmids containing codon-optimized, mutant of the constructs described in A transiently transfected in 293 cells with monoclonal antibody. Constructs containing two inactivating mutations are labeled as 205, 230, single mutants are labeled as 205 or 230. **D.** Western blot analysis of the same samples described in A with polyclonal anti-ORF2p antibodies. **E.** Western blot analysis of the same samples described in B with anti-human ORF2p polyclonal antibodies. GAPDH is used as a loading control. 50-150 kDa and 37 kDa on the right indicate molecular markers. Arrows denote bands of expected molecular weights for each construct listed.

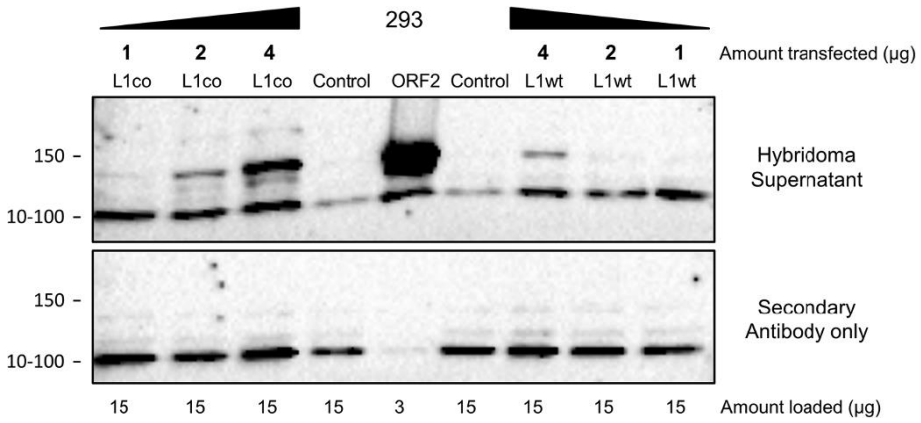

Figure 5. Analysis of ORF2p generated from a functional wild-type and a functional codon-optimized full-length L1 expression plasmids in 293 cells. (top) Western blot analysis of ORF2p generated from expression plasmids containing a full-length wild-type L1 (L1wt), a full-length codon-optimized L1 (L1co) or a codon-optimized ORF2 transiently transfected in 293 cells with supernatant collected from cultured hybridoma cells producing anti-ORF2 antibody. 293 cells were transfected with 1, 2 or 4 micrograms (µg) of the L1wt or L1co expression plasmids or 2 micrograms of the ORF2 expression plasmid and total protein was harvested 24 hours after transfection. Control lane indicates cells transiently transfected with an empty vector. Positions of molecular markers are indicated on the right as 100 or 150 kDa. **(bottom)** The same experiment and analysis as in (top), but using secondary antibodies only. Positions of molecular markers are indicated on the right as 100 or 150 kDa. Total amount of 293 cell lysate loaded is in micrograms (µg).

2.4.5: Monoclonal anti-ORF2p antibody inhibits L1 endonuclease activity in an *in vitro* endonuclease cleavage assay

The unique epitope location within the L1 EN as well as the antibody's ability to detect natively folded ORF2p endonuclease purified from bacterial cells (Supplemental Figure 3) open a possibility that the monoclonal antibody may inhibit L1 endonuclease activity. For this purpose, a previously reported *in vitro* endonuclease cleavage assay was used to measure L1 EN activity (*75, 76*). Figure 7A shows a schematic of the DNA products expected to be observed upon cleavage of the substrate DNA by the L1 EN at the L1 EN site present in the template DNA sequence. Figure 7B demonstrates detection of the expected cleavage products resolved by PAGE when a bacterially purified, functional human L1 EN protein is present in the reaction. The addition of increasing amounts of the monoclonal anti-ORF2p antibody resulted in about 25% reduction of the cleaved products (Figure 7B). This effect was not observed when an unrelated, anti-ORF1p antibody was included in the reaction (Figure 7C). A similar *in vitro* endonuclease cleavage assay using a functionally-related recombinant human apurinic/apyrimidinic endonuclease 1 (APE 1), which shares sequence homology with the L1-encoded endonuclease, was used to test the specificity of this effect. We observed no change in the APE 1 activity upon the addition of the highest amount of the L1 EN-specific antibody (200 nM, Supplementary Figure 4).

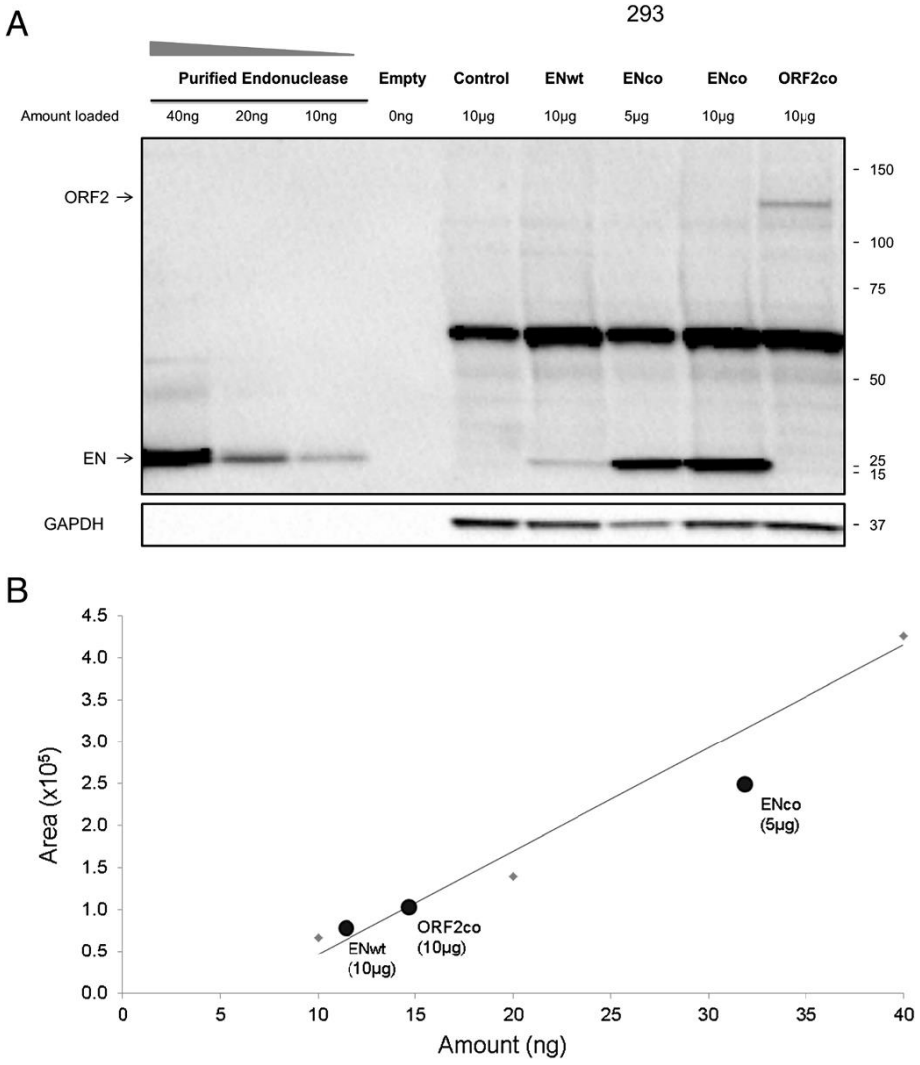

Figure 6. Analysis of sensitivity of the custom anti-human ORF2p monoclonal antibody. A. Western blot analysis of protein generated from expression plasmids containing wild-type ORF2 endonuclease sequence (ENwt), codon-optimized ORF2 endonuclease sequence (ENco) and codon-optimized ORF2 sequence (ORF2) transiently transfected in 293 cells with our monoclonal antibody. 5 or 10 micrograms (µg) of the whole cell lysate was used for analysis as indicated. Control lane indicates cells transiently transfected with an empty vector. Bacterially purified endonuclease was loaded at 0 (empty, buffer only), 10, 20, and 40 nanograms (ng). GAPDH is used as a loading control. 15-150 kDa on the right indicate positions of molecular markers. Arrows denote bands of expected sizes for each protein. **B.** A standard curve was generated using the quantitation of the increasing amounts of the bacterially purified endonuclease shown in A. Signals detected for ORF2co, ENco and ENwt are plotted and labeled with the respective names of the proteins.

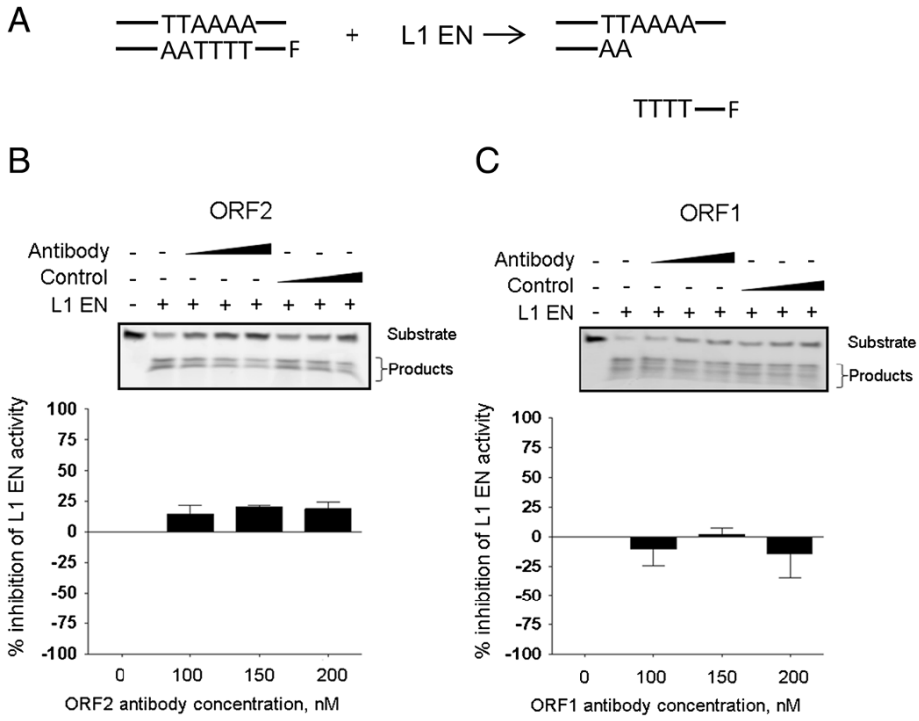

Figure 7. Monoclonal anti-ORF2p antibody inhibits L1 endonuclease activity in an *in vitro* endonuclease cleavage assay. A. Schematic of *in vitro* endonuclease cleavage assay. Double-stranded DNA containing L1 ORF2 endonuclease consensus target sequence with 5' tagged with fluorophore. L1 ORF2 endonuclease is added, DNA is cleaved releasing the fluorphore, which can be quantitated. **B. (top)** SDS-PAGE analysis of the products resulting from the *in vitro* endonuclease assay with or without the addition of the monoclonal anti-ORF2 antibody (0, 100, 150, 200 nM). Antibody (ORF2) denotes the addition of the monoclonal anti-ORF2 antibody, control indicates the addition of the same volume of the buffer used for the reactions containing monoclonal anti-ORF2p antibody, L1 EN denotes bacterially purified human ORF2 endonuclease. **(bottom)** Quantitation of the results of the *in vitro* endonuclease cleavage assay in A (see materials and methods). Results were normalized to 0nM control (N=3). **C.** Same experimental approach as in B, but anti-hORF1p antibody was added to the *in vitro* endonuclease cleavage assay.

2.5: DISCUSSION

L1 is responsible for all of the retrotransposon-induced genomic instability in the human genome, as it is the only active source of the functional ORF1 and ORF2 proteins required for mobilization of LINEs, SINEs, and SVA elements (*2, 85-89*). L1 expression and retrotransposition are suppressed by many diverse cellular pathways, in order to minimize the genomic damage inflicted by L1 activity. (*48, 49, 77, 90-95*). L1 encodes an ORF2 protein with several identified functions essential for the retrotransposition process. These include the endonuclease (*67*) and reverse transcriptase (*68*) activities, and a putative RNA binding domain within the C-terminus of the protein (*70*). Studies geared toward understanding the biological relevance of this multifunctional protein and its effect on human health necessitate manipulations involving changes in protein sequence as well as characterization of the expression of resulting ORF2p variants *in vitro*, in cultured cells, and *in vivo*. To satisfy this requirement, polyclonal antibodies against the L1 ORF2 protein of mouse and human origin have been previously reported (*60, 72, 73*).

We have developed a custom monoclonal antibody to the human L1 ORF2p endonuclease domain that will help to advance future studies involving ORF2p expression and function. The monoclonal nature of the antibody provides a continuous source of antibody, thereby eliminating the inherent issue with reproducibility commonly associated with different batches of polyclonal antibodies raised against the same antigen. Similar to previously reported polyclonal antibodies (*72*) our monoclonal anti-ORF2p antibody detects untagged ORF2 protein expressed from the plasmids containing full-length wild-type or codon-optimized L1 elements. This characteristic is beneficial

because the addition of different tags can interfere with L1 protein function or subcellular localization (*9, 49*). Using bacterially purified endonuclease protein, we generated a standard curve that allowed us to determine the sensitivity of our monoclonal antibodies which is about 10ng of the purified protein under described detection conditions (Figure 6). Consistent with the previous reports, we confirmed that codon-optimization of the human L1 ORF2 sequence results in a 5-6-fold increase in the EN protein production (*78, 84*). As we were unable to detect endogenously expressed L1 ORF2p in HeLa and 293 cells, our calculations suggest that endogenous levels of L1 ORF2p expression are less than 10ng of protein per 10µg of cellular lysate.

Additionally, we determined that our antibody specifically recognizes human, but not mouse, ORF2 protein despite the relatively strong sequence conservation between the endonuclease domains of the two proteins (Supplemental Figure 5) (*78, 81*). This feature is useful for studies involving mouse cells and human ORF2 protein. We also ascertained that the epitope recognized by the monoclonal anti-ORF2p antibody includes amino acid 205 of the human ORF2p endonuclease domain. This amino acid is required for the ORF2p endonuclease activity and is therefore necessary for L1-driven retrotransposition (Figures 3 and 4). As a result our antibody exhibits some bias toward detection of the ORF2 proteins containing a functional endonuclease domain at least relative to the status of amino acid 205. Additionally, an alignment of the consensus L1PA1-PA8 ORF2p sequences (*81*) demonstrated that L1PA3-5 have the same sequence as L1PA1, L1PA 2, 6, 7 have one substitution in the core region surrounding amino acid 205 (200-210aa), and L1PA8 varies by two amino acids from the L1PA1 sequence (Supplemental Figure 5).

This discovery opens up the possibility that our monoclonal antibody may inhibit human L1 endonuclease activity. Suppression of L1 retrotransposition by HIV reverse transcriptase inhibitors has been previously reported (*96, 97*) generating an interest in developing L1-specific inhibitors with the potential to suppress L1-associated damage *in vivo*. While the use of such RT inhibitors serves as a helpful tool to study the L1 replication cycle, these inhibitors are not specific to L1 as they are also expected to suppress telomerase RT (*98*). Furthermore, they have significant side effects in humans (*99*) and it is not known whether the inhibition of the reverse transcriptase also prevents damage from the L1 endonuclease-induced DNA double-stranded breaks (DSBs). Thus, inhibition of L1 endonuclease activity is an attractive approach in order to suppress most, if not all, of the L1-induced damage. The development of either chemical or antibody-based inhibitors are the two of the main approaches generally used for the suppression of enzymatic activities. In addition to the effective inhibition of enzyme activity, efficient delivery, stability, and lack of toxicity are common goals for both types of inhibitors (*100, 101*). The specificity of inhibition is a potential challenge with the development of L1 endonuclease inhibitors because this endonuclease is related to the apurinic/apyrimidinic endonuclease (human APE1), which is involved in the repair of DNA damage by the base excision repair (BER) pathway (*80*). Using a fluorescence-based *in vitro* cleavage assay, we demonstrated that our monoclonal anti-ORF2p antibody can reduce L1 endonuclease activity by about 25% without any inhibitory effect on the *in vitro* activity of the human APE1 (Supplemental Figure 4). While it is not known yet whether the antibody is able to inhibit L1 endonuclease activity in the context of the full-length ORF2 protein or in the cellular environment, these results provide the first proof of

principle that the activity of L1 endonuclease can be reduced by an antibody specific to its position 205.

2.6: CONCLUSIONS

Our data demonstrate that this anti-ORF2p monoclonal antibody will be a useful tool for studies involving human L1 because it is specific to human ORF2p. The anti-ORF2p monoclonal antibody detects ORF2 protein generated from the ORF2 expression plasmid as well as both codon-optimized and wild-type full-length L1 expression plasmids transiently transfected into human cells. Our data also establish a rationale for the development of antibody-based inhibitors of L1-induced damage.

CHAPTER 3: CHARACTERIZATION OF L1 ORF1P SELF-INTERACTION AND CELLULAR LOCALIZATION USING A MAMMALIAN TWO-HYBRID SYSTEM

Published as: Mark Sokolowski, Dawn deHaro, Claiborne M. Christian, Kristine J. Kines, Victoria P. Belancio. Characterization of L1 ORF1p self-interaction and cellular localization using a mammalian two-hybrid system. *PLoS ONE*. 2013;8(12):e82021. doi: 10.1371/journal.pone.0082021.

3.1: ABSTRACT

Long INterspersed Element-1 (LINE-1, L1) is an active retrotransposon that mobilizes using a ribonucleoprotein particle (RNP) intermediate composed of the full-length bicistronic L1 mRNA and the two proteins (ORF1p and ORF2p) encoded by that mRNA. ORF1p and ORF2p demonstrate *cis*-preference for their encoding mRNA. Previous studies of ORF1p, purified from bacterial and insect cells demonstrated that this protein forms trimers *in vitro*. While valuable for understanding ORF1p function, these *in vitro* approaches do not provide any information on ORF1p self-interaction in the context of mammalian cells. We used a mammalian two-hybrid (M2H) system in order to study L1 ORF1p self-interaction in human and mouse cells. We demonstrate that the M2H system successfully detects human and mouse ORF1p self-interactions in transiently transfected mammalian cells. We also generated mouse and human ORF1p-specific antibodies to characterize the expression of ORF1p fusion proteins used in the M2H system. Using these antibodies, we demonstrate that ORF1p interaction *in trans* leads to

the formation of heterodimers that are expected to produce a positive signal in the M2H system. Although the role for L1 ORF1p *cis*-preference in L1 mobilization is established, the impact of ability of ORF1pto interact *in trans* on the L1 replication cycle is not known. Furthermore, western blot analysis of ORF1p generated by a full-length L1, wild type ORF1, or a codon-optimized ORF1 expression vector is detected in the nucleus. In contrast, the addition of a tag to the N-terminus of the mouse and human ORF1 proteins can significantly alter the subcellular localization in a tag-specific manner. These data support that nuclear localization of ORF1p may contribute to L1 (and potentially the SINE Alu) RNP nuclear access in the host cell.

3.2: INTRODUCTION

Long INterspersed Element-1 (LINE-1, L1), an autonomous non-long terminal repeat retrotransposon, has contributed greatly to the evolution of the human genome through retrotransposition of itself and facilitation of retrotransposition of the parasitic SINE Alu and SVA elements (*7, 14, 62*). There are roughly 500,000 related copies of L1 distributed throughout the human genome (*10*). Though the majority of these loci are 5'-truncated, there are 80-100 full-length L1 copies that are predicted to retain activity (*10, 24-27*). A full-length, autonomous L1 is composed of a 5' untranslated region (UTR) with an internal promoter, two open reading frames (ORF1 and ORF2), and a 3' UTR ending in a polyA site and associated polyA tail (*19, 20*). ORF1 and ORF2 proteins (ORF1p and ORF2p) are translated from the bicistronic L1 mRNA (*21*) and, potentially, from prematurely polyadenylated and spliced L1 mRNAs (*77, 93*). Association of ORF1p, ORF2p, and the full-length L1 mRNA which generated these proteins into a

ribonucleoprotein particle (RNP) is required for L1 retrotransposition (*7, 29*). The protein components of the L1 RNP exhibit *cis*-preference for their encoding L1 mRNA, thereby limiting the retrotransposition of cellular mRNA and L1 mRNA produced by defective L1 loci (*5, 36*).

ORF1p is a 40 kDa protein with an N-terminal domain, coiled-coil domain (CCD), an RNA recognition motif (RRM) and a C-terminal domain (CTD) (*38*). ORF1p exhibits RNA binding and nucleic acid chaperone properties, which are required for L1 mobilization (*29, 102*). Human and mouse ORF1p (hORF1p and mORF1p,) trimerize through their respective CCDs and associate with their corresponding full-length mRNAs, a process that is crucial for L1 RNP formation (*3, 37, 39, 41*). It is estimated that an ORF1p trimer occupies about 50 nucleotides on the template RNA, suggesting that ORF1p trimers may be the most abundant component in the L1 RNP (*3, 37, 41*).

As part of the L1 replication cycle, the L1 RNP must enter the nucleus. While this is an essential step for L1 amplification, the mechanism of L1 nuclear entry is currently unknown. It has been suggested that L1 ORF2p plays a role in the nuclear localization of the L1 RNP (*73*), while any contribution of L1 ORF1p to the subcellular localization of the L1 RNP is poorly understood. ORF1p fused to a green fluorescent protein (GFP), as well as other tags, has been detected primarily in the cytoplasm by immunohistochemistry (IHC) (*103*). The cytoplasmic localization of this fusion protein was affected by various truncations of the ORF1p portion of the chimeric protein (*103*). Recently, endogenously expressed ORF1p has been detected by IHC in both the nuclei and cytoplasm of cells of human tumor samples collected from patients with different

disease stages (*54*, *55*). Nuclear and cytoplasmic L1 ORF1p was also reported to interact with multiple cellular proteins in both compartments (*48*). These data demonstrate that the subcellular localization of ORF1p may vary depending on the cell type, and suggest the possibility that its localization may be affected by the addition of tags. While these examinations of ORF1p localization are, by necessity, carried out *in vivo,* studies dealing with ORF1p self-interaction have almost exclusively been performed *in vitro*.

The majority of the critical findings regarding ORF1p trimerization have been made *in vitro* using human or mouse proteins purified from *E. coli* or *baculovirus*-infected insect cells (*37, 39, 41, 102*). The ORF1p behavior *in vitro* is presumed to translate to the mammalian environment. However, limited experimental information exists concerning the properties and functions of ORF1p in mammalian cells. While the data acquired *in vitro* are invaluable, these approaches are also laborious, technically challenging, and do not account for the potential influence of host cellular factors on ORF1p self-interaction. Recently, many cellular proteins have been reported to interact with L1 ORF1p in the nucleus and cytoplasm (*48, 104*), suggesting that cellular factors, unaccounted for in *in vitro* approaches, may influence the behavior of ORF1p. To capture L1 ORF1p self-interaction in a more biologically relevant manner, we used the mammalian two-hybrid system (M2H). A major advantage of utilizing this approach is that it recapitulates some aspects of ORF1p biology related to L1 replication cycle such as ORF1p self-interaction. From here on, the ORF1p self-interaction is defined as dimerization or trimerization of ORF1p molecules translated from the same or different mRNAs.

Our data demonstrate that the M2H system detects human and mouse ORF1p self-interactions in transiently transfected human and mouse cells. Using polyclonal antibodies specific to human and mouse ORF1p, we demonstrate that the interactions detected by the M2H system are due, at least in part to the formation of heterodimers *in trans*. ORF1p heterodimerization *in trans* or self-interaction *in trans* is defined as the interaction between ORF1 proteins made from different mRNAs. ORF1p heterodimerization occurs with various efficiencies in human and mouse cells as detected by western blot analysis and the M2H system. We also detected significant presence of ORF1p in the nucleus and demonstrate that the fusion of various tags to the N-terminus of the ORF1p alters its localization in a tag-specific manner. Nuclear localization of the untagged L1 ORF1p expressed from the full-length wild type (wt) L1 supports the theory that L1 ORF1p may assist L1 RNPs in gaining access to the nucleus.

3.3: MATERIALS AND METHODS

3.3.1: Cells

NIH-3T3 (ATCC CRL-1658) and HeLa (ATCC CCL2) cells were maintained as previously described (*77*). PC3 cells (ATCC CRL-1435) were cultured in HyClone Dulbecco's modified Eagle's medium (DMEM) with L-Glutamine and Sodium Pyruvate (Thermo Scientific) with 10% fetal bovine serum (Invitrogen) and maintained under 5% CO_2 at 37°C.

3.3.2: Transfections

Western blot: HeLa and NIH-3T3 cells were seeded at $2x10^6$ and $2.5x10^6$ cells per T75 flask, respectively, and transfected 16-18 hours later with 5µg of expression plasmids containing wt and 3µg of plasmids containing codon-optimized mouse or human L1 sequences. Plus reagent (6µl) (Invitrogen) and lipofectamine (18µl (3T3) or 24µl (HeLa)) (Invitrogen) were used in the transfection reaction in serum-free media. Serum-free media was replaced with serum-containing media 3 hours after transfection and cells were harvested 24 hours later. Dose curve of ORF1p expression was performed by transfecting 1-6 micrograms of hORF1co expression plasmid, using the same conditions as above. Equal loading western blot: (Supplemental Figure 4B). The loading amount of the western blot was the only deviation from the above mentioned western blot protocol. The amounts for each nuclear and cytoplasmic compartment pair were determined by the compartment that had the lowest percentage of total protein. The relative percentage of each compartment was between 2.1-3.8% of the total protein present in that compartment. Mammalian Two-hybrid (M2H, Promega): HeLa and 3T3 cells were

seeded at $1.5x10^5$ and $1x10^5$ cells per well in a 6 well plate 18 hours prior to transfection. Each of the three M2H plasmids (0.1µg) were cotransfected using 1µl of plus reagent (Invitrogen) and 1.5µl of lipofectamine (Invitrogen). Media containing transfection cocktail was changed 3 hours after transfection and cells were harvested 48 hours later.

3.3.3: Protein harvest for western blot analysis

Cells were harvested using Triton lysis buffer, TLB, (50 mM Tris, 150mM NaCl, 10 mM EDTA, 0.5% TritionX 0.5%, pH=7.2) for collection of cytoplasmic protein or Triton sodium dodecyl sulfate lysis buffer, TLB SDS, (50 mM Tris, 150mM NaCl, 10 mM EDTA, 0.5% sodium dodecyl sulfate, TritionX 0.5%, pH=7.2) for nuclear and total protein harvesting. Lysis buffers were supplemented with Halt Protease inhibitors, and phosphatase inhibitors 2 and 3 (Sigma) at 10µl/mL each. The total and nuclear lysates were sonicated three times for 10 seconds at 12 watts RMS using a 3mm wide QSonica Microson homogenizer with Microson ultrasonic cell disruptor XL2000 (Microson).The protein concentrations of the lysates were determined using 595nm wavelength OD values against a Bovine Serum Albumin (BSA) standard.

For western blot analysis, samples (10-40µg) were combined with 2× Laemmli buffer, 1.6µl β-mercaptoethanol and boiled for 5 minutes prior to loading. Samples were fractionated on Bis-Tris 4-12% Midi gels (Invitrogen) and transferred to nitrocellulose membranes (iBlot System; Invitrogen). Membranes were blocked for 1 hour in 5% milk in PBS-Tween and incubated with hORF1 (custom polyclonal Rabbit antibodies, epitope: TGNSKTQSASPPPK), mORF1 (custom polyclonal Goat antibodies, epitope: YRTPNRLDQKRNSS) or T7 (Cell signaling 6885s) antibodies overnight at 4°C.

Detection was carried out using horseradish peroxidase-conjugated secondary antibodies (Santa Cruz sc-2317) and developed using Immun-Star™ WesternC™ Kit (Bio-Rad, Cat. #170-5070). Equal loading and the quality of subcellular fractionation was confirmed with GAPDH antibodies (Santa Cruz sc-25778), Lamin A/C (Cell signaling 2032S) or Calregulin (Santa Cruz T-19) using the same protocol.

Non-reducing gels (Figure 5 and Supplemental Figures 5, 6, 7 and 8). Transfected cells were lysed using TLB buffer supplemented with 0, 1, 13, or 25mM of N-ethylmaleimide (NEM) (Sigma E3876) dissolved in 100% ethanol (Sigma E7023). Non-NEM samples were supplemented with 100% ethanol equal in volume to the added NEM. The samples were sonicated 5 times for 5 seconds and kept on ice between sonications. Samples were combined with 2× Laemmli buffer without β-mercaptoethanol and heated at 85°for 5 minutes. Samples were fractionated on non-reducing denaturating SDS -PAGE (Bis-Tris 4-12% Midi gels (Invitrogen)) gels. Unless otherwise indicated, 25 mM of NEM was used.

Heterodimer formation analysis HeLa cells were individually transfected with ORF1 expression plasmids. Cells were transiently transfected with hORF1 or hGORF1 expression plasmids, trypsinized 24 hours later and mixed in equal amounts (supplemental figure 6). Total protein from the cell mixture was harvested as above with or without NEM. Protein samples were harvested without NEM from NIH-3T3 cells transiently transfected with hORF1 or hGORF1, then mixed in 1:1 ratio and incubated for 0, 5, 10, or 15 minutes (supplemental figure 7). Total protein was harvested with or without NEM from NIH-3T3 cells cotransfected with hORF1 and hGORF1 expression

plasmids. Cells were then exposed to additional sonication in the presence of absence of NEM (supplemental figure 8).

3.3.4: Mammalian Two-Hybrid

Cells were harvested by scraping into 250 µL of 1x passive lysis buffer (Promega, E1960) subjected to freeze(-80°C)/thaw(25°C) lysis x3. Supernatant was collected after centrifugation for 30 seconds at $1.61x10^4$rpm at 4°C. The protein was quantitated using a BSA assay as described above. The cellular lysates were assayed according to manufacturer's protocol (Promega E1960, E2440) using Sirius luminometer (Titertek Berthold, Pforzheim, Germany).

3.3.5: Plasmids

"hL1wt" is JM101/L1.3 no tag (L1.3) (*36*). "hORF1co" is pBudORF1opt (*15*). "mORF1co" is pBudORF1$_{syn}$ (*84*). "mL1wt" is L1$_{Spa}$ (*82*). "hORF1wt" is pBudORF1 L1.3 (*15*). Human and mouse wild type L1 sequences used for cloning are L1.2 (L19088) and L1spa (AF160099) elements. pBud and pBudCE4.1 (Invitrogen) were used as the empty vector controls. hGORF1co expression plasmid was generated by PCR amplification of "hORF1co" with primers containing SgfI and PmeI restriction sites and subcloning the PCR fragment into the pFN11A vector (Promega). hVORF1co was generated by PCR amplification of "hORF1co" with primers containing SgfI and PmeI restriction sites and subcloning the PCR fragment into the pFN10A vector (Promega). The same approach using the above-mentioned expression vectors was used to generate hGORF1wt, hVORF1wt, mGORF1co, mVORF1co, mGORF1wt, mVORF1wt expression plasmids.

3.3.6: Statistical Analysis

Statistical significance was determined using a one-tailed T-test with equal variance in excel using N=3 to 5. Error bars represent standard deviation.

3.4: RESULTS

3.4.1: Mammalian two-hybrid system detects human and mouse ORF1p self-interaction

We used the mammalian two-hybrid system (M2H) to assess ORF1p self-interaction in transiently transfected mammalian cells. The M2H system utilizes three expression vectors that allow for testing interactions between two specific proteins (Figure 1A) which are fused to either a GAL4 DNA-binding domain or a VP16 transcriptional activation domain. The GAL4 binding domain of the interacting pair is expected to bring cellular transcription machinery recruited by the VP16 activation domain to a reporter plasmid containing GAL4 binding sites upstream of a promoterless firefly luciferase gene. Therefore, luciferase activity is detected only if the GAL4- and the VP16-fused proteins stably interact in mammalian cells (Figure 1A). The positive control used in both HeLa and 3T3 cell lines is the interaction between MyoD, a myogenic regulatory protein, and Id, a negative regulator of myogenic differentiation. The negative control is GAL4-fused ORF1 codon-optimized (GORF1co) cotransfected with untagged ORF1co. Both human and mouse ORF1p trimerize and multimerize *in vitro* (*37, 39*). If these interactions are preserved in the mammalian environment, the M2H system is expected to detect interactions between ORF1 GAL4 and VP16 homotrimers or monomers, provided that ORF1p can interact *in trans*. ORF1p interactions *in trans* are defined as interactions between two or more ORF1 proteins generated from different mRNAs.

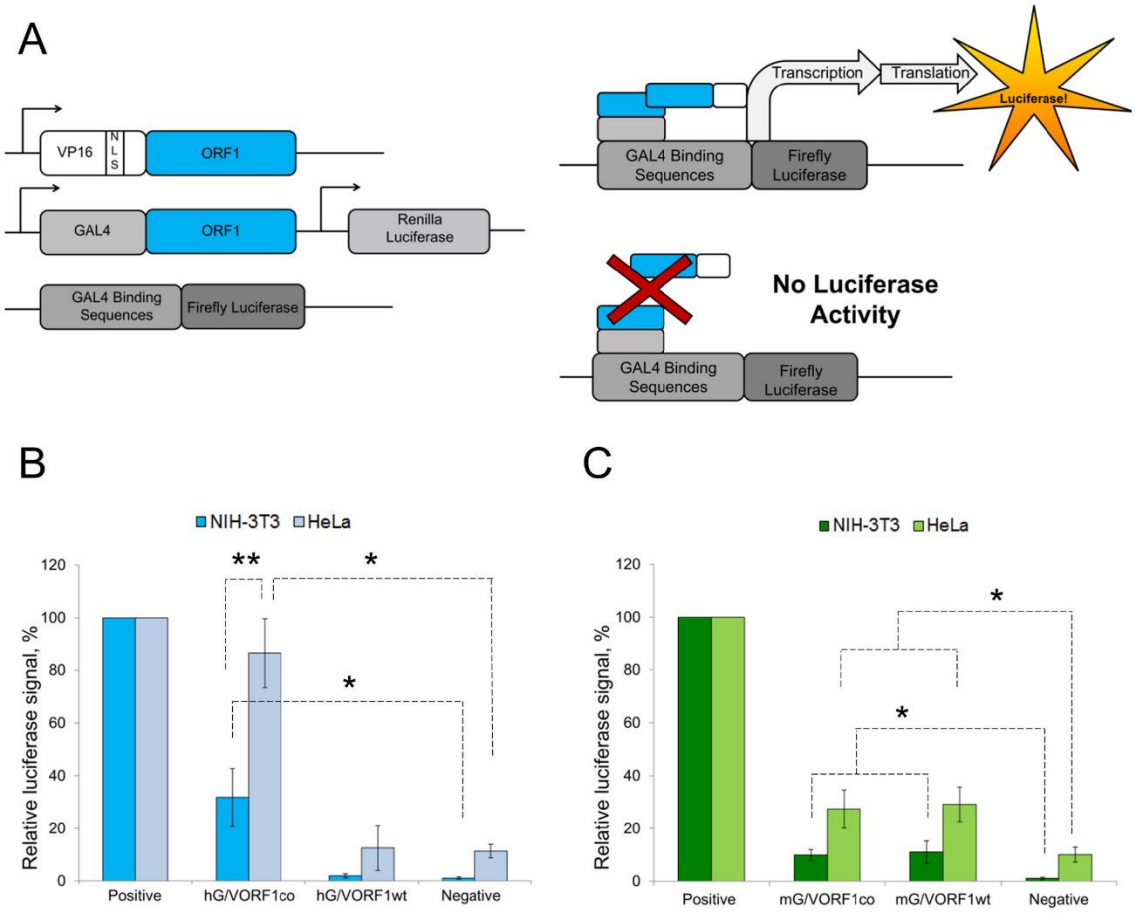

Figure 1. Analysis of ORF1p interaction using the M2H system. A. (left) Schematic of the M2H constructs, arrows denote promoters. One ORF1p is fused to a GAL4 binding protein, derived from the GAL4 yeast transcription factor which binds a specific sequence of DNA. The other ORF1 protein is fused to the VP16 viral transcriptional transactivator protein sequence, with an added SV40-derived internal nuclear localization signal. Tandemly-arrayed GAL4 binding sites are present upstream of the firefly luciferase gene. **(top right)** Interaction between proteins of interest expressed from the M2H expression plasmids leads to firefly luciferase expression. **(bottom right)** If the proteins do not interact with one another, no firefly luciferase is produced. **B.** M2H results for codon-optimized and wild type hGORF1p:hVORF1p interaction in HeLa and NIH-3T3 cells normalized to the positive control. Asterisk denotes statistical significance between the relative luciferase signals detected for hG:VORF1p interaction (wt or co) in HeLa or NIH-3T3 cells compared to their respective negative control (p-value <0.05). Double asterisk denotes statistical significance between relative signal reflecting interaction between hORF1co and VORF1co detected in HeLa versus NIH-3T3 cells (p-value<0.05). **C.** M2H results for codon-optimized and wild type mGORF1p:mVORF1p interactions in HeLa and NIH-3T3 cells normalized to their respective positive controls. Asterisk denotes statistical significance between the relative luciferase signals detected for mG:VORF1p (wt or co) when compared to their respective negative controls (p-value<0.05).

We assessed human ORF1p self-interactions using the M2H system by subcloning codon-optimized ORF1 sequences (*15, 84*) into the GAL4 and VP16 expression vectors (see material and methods and Supplemental Table 1). Transient transfection of HeLa cells with human codon-optimized GAL4- and VP16-ORF1 fusion expression plasmids (hGORF1co and hVORF1co) detected robust ORF1p self-interaction (Figure 1B). Weaker human ORF1p self-interaction was detected in mouse NIH-3T3 cells as compared to human cells (Figure 1B). Analysis of luciferase activity in cells transiently transfected with GAL4- and VP16-fused codon-optimized mouse ORF1 expression plasmids (*84*) (mGORF1co and mVORF1co, respectively) demonstrated that HeLa cells supported more efficient self-interaction of mouse ORF1p than did NIH-3T3 cells (Figure 1C, mG/VORFco). In all experiments, Renilla luciferase generated from the GORF1 expressing plasmid was used to normalize for transfection efficiency.

To determine whether self-interaction of human and mouse ORF1 proteins produced by wild type (hORF1wt and mORF1wt) rather than codon-optimized L1 ORF1 sequences can be detected by the M2H system, we subcloned wt human and wt mouse ORF1 sequences into the M2H expression vectors. Figure 1C demonstrates that mouse ORF1p produced from the wild type ORF1 sequence can self-interact as efficiently as the protein generated from the plasmid containing codon-optimized ORF1 sequence when transiently transfected in HeLa cells (Figure 1C, mG/VORF1wt and mG/VORF1wt). However, there was no self-interaction detected for the human ORF1p generated from the wt expression plasmid in either cell line (Figure 1B).

We also tested the possibility that mouse and human ORF1p may interact with one another in mammalian cells. Cotransfection of M2H plasmids expressing hORF1co and mORF1co fusion proteins in HeLa and NIH-3T3 cells did not produce any luciferase signal, demonstrating the species specificity of ORF1p interactions (Supplemental Figure 1). These data demonstrate that the M2H system successfully detects specific ORF1p self-interactions in mammalian cells. This system complements existing *in vitro* methods and provides a unique alternative for testing ORF1p interactions.

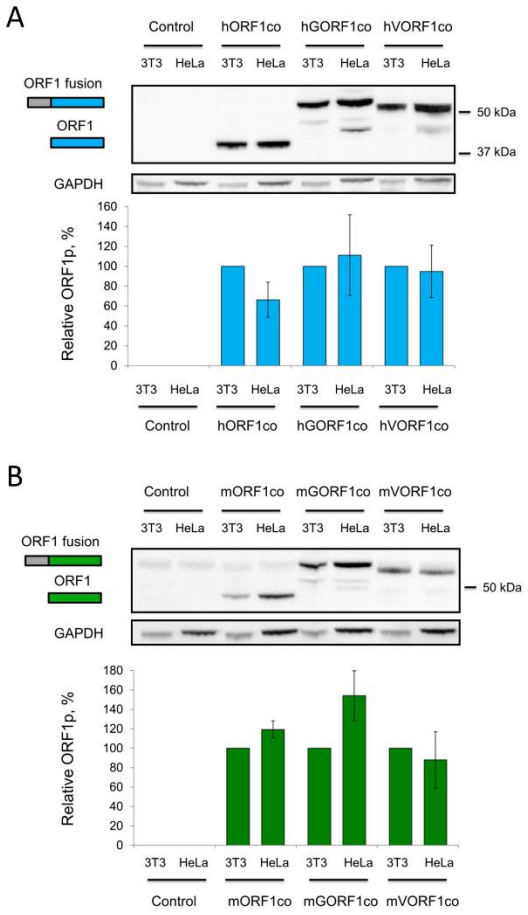

Figure 2. Analysis of M2H ORF1 fusion protein expression in human and mouse cells. A. (top) Western blot analysis of codon-optimized human ORF1 (hORF1co, 40 kDa, blue rectangle labeled as ORF1) and human ORF1co N-terminally fused to GAL4 or VP16, (hGORF1co, 57 kDa, and hVORF1co, 49 kDa, blue and grey rectangles labeled as ORF1 fusion) transiently transfected in HeLa or NIH-3T3 cells. Human ORF1 protein was detected with custom-made human-specific ORF1 polyclonal antibodies. GAPDH is used as a loading control. 37 and 50 kDa are molecular markers. Control lanes indicate cells transfected with an empty vector. **(bottom)** Quantitation of western blot results. Signals obtained for hORF1p and its fusions were normalized to their respective GAPDH loading controls and expressed as a percentage of the relative signal detected for each protein in NIH-3T3 cells. **B. (top)** Western blot analysis of codon-optimized mouse ORF1 (mORF1co, 47 kDa, T7-Myc-His tag, green rectangle labeled as ORF1) and mouse ORF1co N-terminally fused to GAL4 and VP16 (mGORF1co, 60 kDa, and mVORF1co, 52 kDa, green and grey rectangles labeled as ORF1 fusion) transiently expressed in HeLa or NIH-3T3 cells. 50 kDa is a molecular marker **(bottom)** Quantitation was performed as described in A.

3.4.2: ORF1 protein expression in human and mouse cells

The difference in relative ORF1p self-interactions observed between human and mouse cells as well as the lack of detectable self-interaction between the human ORF1 fusion proteins generated from wt plasmids prompted the investigation of expression of the ORF1 fusion proteins in these cell lines. To accomplish this analysis, we generated human and mouse L1 ORF1p-specific antibodies (Supplemental Figure 2A and 2B). Western blot analysis of human ORF1p produced by the plasmids containing codon-optimized ORF1 sequence (hORF1co, expected size 40 kDa), GAL4- (hGORF1co, expected size 57 kDa) or VP16-ORF1co (hVORF1co, expected size 49 kDa) sequences demonstrated similar steady-state levels of total protein for each of these constructs in both human and mouse cells (Figure 2A). A similar result was observed for the respective mouse proteins, mORF1co (expected size 47 kDa due to a T7-Myc-His tag), mGORF1co (expected size 60 kDa), mVORF1co (expected size 52 kDa) (Figure 2B), demonstrating that the overall protein expression did not account for the differences detected by the M2H system (Figure 1B and 1C).

Western blot analysis of the total ORF1 protein generated by the M2H expression plasmids containing wt human L1 ORF1 sequences, hGORFwt (expected size 57 kDa) and hVORFwt (expected size 49 kDa), did explain the lack of signal from these expression plasmids in the M2H assay (Supplemental Figure 3A). No detectable protein was produced by either one of these plasmids, even when transfection conditions were manipulated in an attempt to increase the signal (data not shown). In contrast, the unfused human ORF1p expressed by the full-length wt L1 (hL1wt, expected size 40 kDa) or

ORF1 expression plasmids (hORF1wt , expected size 41 kDa due to a T7 tag, hORF1co, expected size 40 kDa) was readily detectable in HeLa cells (Supplemental Figure 2A). We speculate that our inability to detect human fusion proteins generated from wt plasmids can be explained by splicing between L1 splice sites and an intron present in the M2H expression vectors (*93, 105, 106*). Western blot analysis of ORF1 protein generated by the wt mouse L1 ORF1 sequences subcloned into the M2H expression vectors detected mouse GAL4 and VP16 fusion proteins in HeLa cells (Supplemental Figure 3B, nuclear fraction). However, the steady-state levels of mVORF1wt were significantly reduced in comparison to the mGORF1wt protein. Western blot analysis of the corresponding mouse proteins demonstrate that mGORF1wt, but not mVORF1wt, is expressed in NIH-3T3 cells (Supplemental Figure 3C), providing an explanation for the difference in the M2H signal between the two cell lines (Figure 1C).

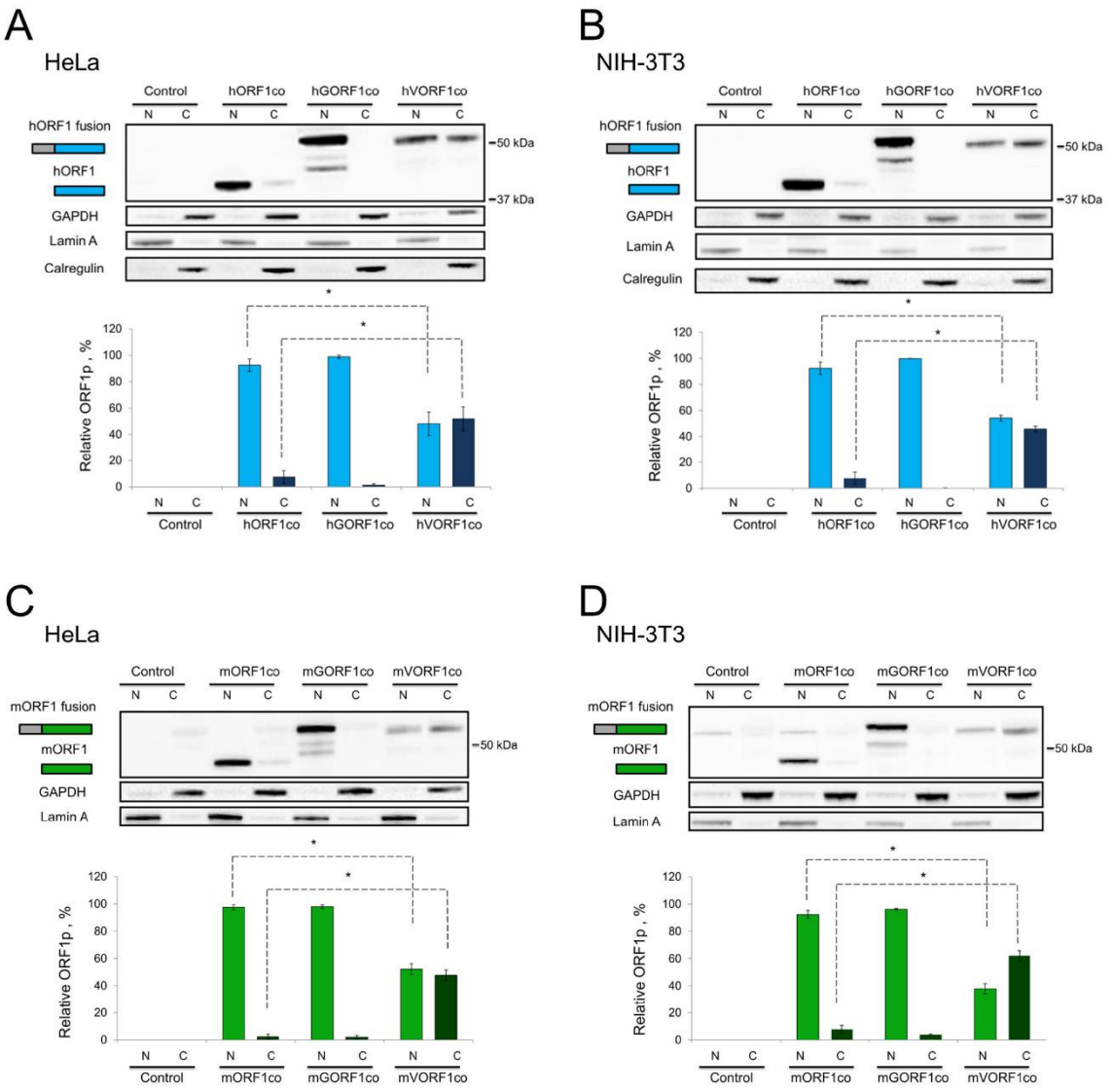

Figure 3. Analysis of subcellular localization of ORF1 protein fusions in human and mouse cells. A. (top) Western blot analysis of subcellular localization of hORF1co (blue rectangle labeled as hORF1), hGORF1co and hVORF1co (grey/blue rectangle labeled as hORF1 fusion) transiently expressed in HeLa cells. Subcellular localization of the human ORF1 protein detected in nuclear (N) and cytoplasmic (C) fractions with human-specific ORF1p polyclonal antibodies. GAPDH (cytoplasmic marker), Lamin A (nuclear marker), and Calregulin (endoplasmic reticulum marker) proteins are used as loading and cell fractionation controls. Control lanes indicate cells transfected with empty vector. **(bottom)** Quantitation of western blot results. hORF1 and its fusions were normalized to their respective GAPDH and Lamin A loading controls. The relative nuclear and cytoplasmic ORF1p was calculated as a fraction of the relative total ORF1p detected in respective cell types. **B.** The same experiment and analysis as in A, but using NIH-3T3 cells. **C.** Western blot analysis as above using mORF1co (green rectangle labeled as mORF1), mGORF1co or mVORF1co (grey/green rectangle labeled as mORF1 fusion) in HeLa cells. Mouse specific ORF1p polyclonal antibodies were used for protein detection. **D.** The same experiment and analysis as in C, using NIH-3T3 cells. For all panels, statistically significant data points are indicated by dashed lines. An asterisk represents a significant difference between the subcellular localization of ORF1p and VP16-fused ORF1p (T-test, p-value <0.05).

3.4.3: Subcellular localization of human and mouse ORF1 proteins

While the total steady-state protein levels of GAL4- and VP16-fused human ORF1 generated from codon-optimized expression plasmids do not differ between NIH-3T3 and HeLa cells, the subcellular localization of these fusion proteins may vary between the cell lines and potentially have an impact on the interactions required to produce a signal in the M2H system. We performed nuclear/cytoplasmic fractionation of HeLa and NIH-3T3 cells transiently transfected with codon-optimized expression plasmids generating untagged, GAL4-, or VP16-fused ORF1p. Western blot analysis demonstrated that untagged hORF1co and GAL4-fused hGORFco are detected predominantly in the nucleus of both cell types (Figure 3A and B). In contrast, a much higher proportion of the VP16-fused ORF1p (*i.e.*, hVORF1co *vs.* hORF1co) is detected in the cytoplasmic fractions of HeLa and NIH-3T3 cells (Figure 3A and B), despite the fact that the VP16 fusion ORF1p contains a nuclear localization signal (NLS). The quantification of western blot results was done by normalization of the ORF1p signal in each fraction to its respective loading controls (GAPDH and Lamin A). The relative nuclear and cytoplasmic ORF1p was calculated as a percent of the combined nuclear and cytoplasmic ORF1p detected in each cell type (Figure 3A and B). Western blot analysis of the respective tagged and untagged mouse ORF1 proteins demonstrated the same pattern of subcellular distribution as was detected for the human ORF1p (Figure 3C and 3D). These results demonstrate that ORF1p subcellular localization can be significantly altered by the addition of an N-terminal tag.

Human ORF1p C-terminally-fused to a green fluorescent protein (ORF1-GFP) has been previously reported to localize predominantly to the cytoplasm when detected by immunohistochemistry (IHC) (*103*). Recently, endogenous L1 ORF1p was detected by IHC in the nuclei of human tumor samples (*54, 55*), suggesting that the presence of a tag or the cell type may influence subcellular localization of the ORF1 protein. Figure 3 demonstrates that human and mouse ORF1 proteins expressed from the plasmids containing codon-optimized L1 ORF1 sequence predominantly localize to the nucleus, as do the GAL4-fused versions of both proteins (hGORF1co and mGORF1co), in both HeLa and NIH-3T3 cells. In contrast, both human and mouse ORF1 proteins fused to VP16 were found in the nucleus and cytoplasm of HeLa and NIH-3T3 cells (Figure 3, hVGORF1co and mVORF1co). To rule out any potential influence of codon-optimization on the subcellular localization of ORF1p, we examined the subcellular localization of ORF1 proteins expressed from plasmids containing wt L1 and ORF1 sequences. Figures 4A and 4B demonstrate that ORF1p expressed by the plasmid containing wild type ORF1 sequence or by the wild type full-length human L1 is found predominantly in the nuclear fraction of human (HeLa and PC3, Supplemental Figure 4A) and mouse (NIH-3T3) cells (Figure 4C). The plasmid containing wt ORF1 sequence expresses an ORF1p that is C-terminally fused to a T7 tag. Our data demonstrate that the T7 tag does not alter ORF1p subcellular localization consistent with the report that this tagged ORF1p supports L1 retrotransposition (*5*). Similarly, mouse ORF1p transiently expressed by the wild type full-length L1 (mL1wt) and mORF1co expression plasmids demonstrated preferential nuclear localization in both human and mouse cells (Figure 4C, mL1wt and mORF1co, respectively).

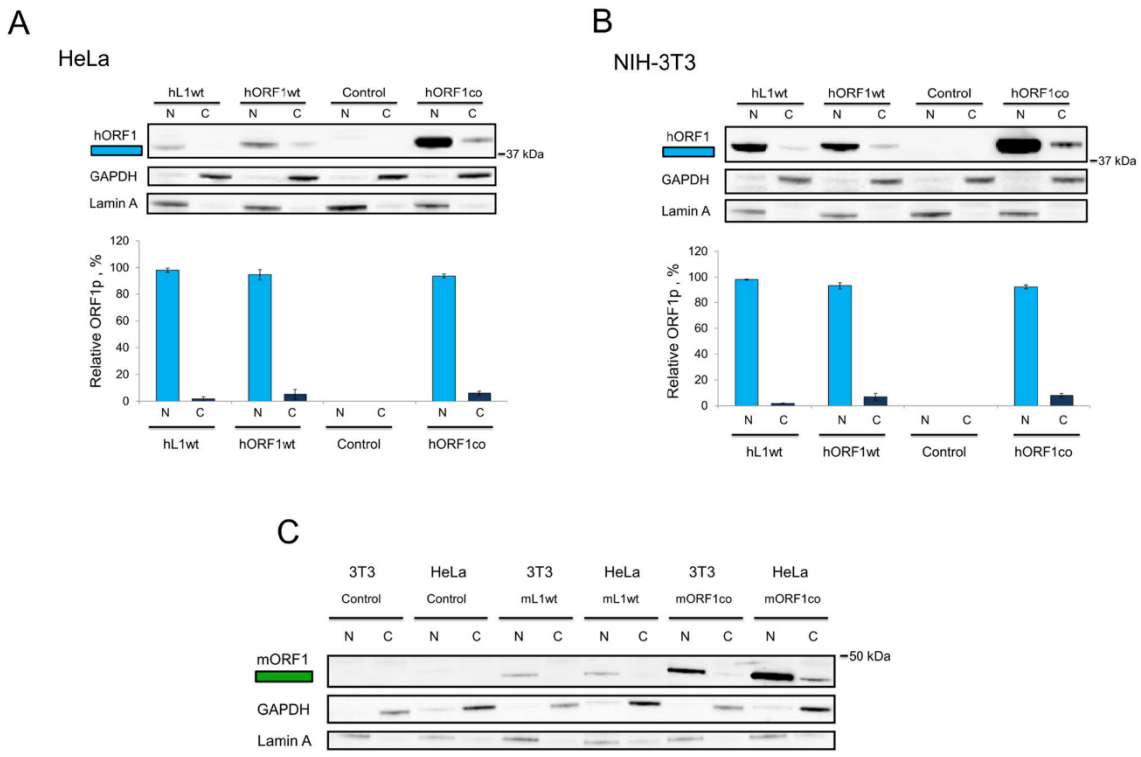

Figure 4. Analysis of subcellular localization of hORF1p expressed from vectors containing wild type sequences. A. (top) Western blot analysis of subcellular localization of ORF1p (blue rectangle labeled as hORF1) expressed from human full-length L1 (hL1wt, 40 kDa), human ORF1 wt (hORF1wt, 41 kDa) or hORF1co (40 kDa) vectors transiently expressed in HeLa cells. Human ORF1p is detected in nuclear (N) and cytoplasmic (C) fractions with human-specific ORF1p polyclonal antibodies. GAPDH (cytoplasmic marker) and Lamin A (nuclear marker) are used as loading and cell fractionation controls. 37 kDa is a molecular marker. Control lanes indicate cells transfected with empty vector. **(bottom)** Quantitation of western blot results: each construct expressing hORF1p was normalized to its respective GAPDH and Lamin A loading controls. The relative amounts of nuclear and cytoplasmic ORF1p were calculated as a fraction of the relative total ORF1p detected. No statistical significance was found between the subcellular localization of ORF1p expressed by the above-described constructs. **B.** The same experiment and analysis as in A was performed in NIH-3T3 cells. No statistical significance was found between the subcellular localization of ORF1p expressed by the above-described constructs. **C.** Subcellular localization of mORF1p (green rectangle labeled as mORF1) transiently expressed from the full-length mouse L1 (mL1wt) or codon-optimized mORF1 (mORF1co) in HeLa or NIH-3T3 cells using mouse specific polyclonal antibodies. 50 kDa is a molecular marker.

Figures 3 and 4 show western blot results obtained by comparing equal amounts of protein collected from nuclear and cytoplasmic fractions. This approach does not take into account the fact that these total amounts represent different percentages of the total protein harvested from each compartment. Loading equal percentages of the total protein harvested from each compartment shows that hORF1p is still predominantly detected in the nuclear fraction of transiently transfected cells (Supplemental Figure 4B).

It is possible that the time elapsed between transfection and protein harvest, as well as the intracellular concentrations of ORF1p, could affect its subcellular localization. ORF1p produced from the full-length human L1.3 expression vector was detected predominately in the nucleus as early as 12 hours after transfection (Supplemental Figure 4C). This finding supports that ORF1p nuclear localization is present as soon as the protein has reached detectable levels. Western blot analysis of nuclear and cytoplasmic fractions from HeLa cells transiently transfected with increasing amounts of codon-optimized human ORF1 expression plasmid further demonstrated that the nuclear localization of L1 ORF1p is consistent within the experimental range (Supplemental Figure 4D).

To further eliminate the influence of methodological artifacts on the ORF1p subcellular localization, we tested different processing and centrifugation conditions for nuclear/cytoplasmic fractionation (Supplemental Figure 4E), all of which resulted in ORF1p detection in the nucleus. These data support our observation that the difference in the subcellular localization of the human and mouse VORF1 proteins, when compared to the unfused ORF1p appears to be tag-specific, as both human and mouse GAL4-ORF1

fusion proteins retain their nuclear localization (Figures 3 and 4). This tag-specific effect on L1 ORF1p localization should be taken into consideration when conclusions about L1 ORF1p biology are drawn from the studies of tagged proteins.

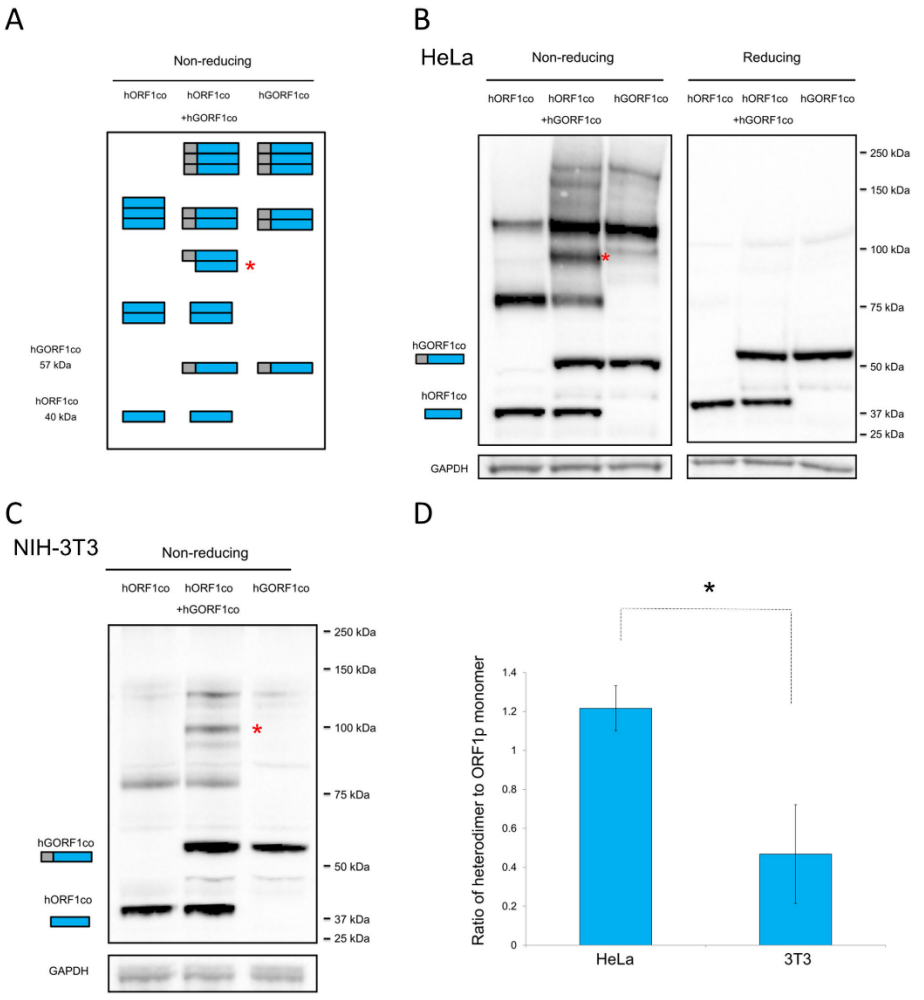

Figure 5. Analysis of ORF1p interactions *in trans*. A. Schematic of predicted ORF1p species expected to be detected if ORF1p heterodimerization takes place when cells are transiently cotransfected with tagged and untagged ORF1p expression plasmids. Blue rectangle represents ORF1, grey rectangle represents a GAL4 tag, with expected molecular weights expressed in kilodaltons (kDa). Red asterisk denotes heterodimers. **B.** Western blot analysis of ORF1p species expressed by hORF1co (blue rectangle labeled hORF1co) and hGORF1co (grey/blue rectangle labeled hGORF1co) plasmids individually transfected (hORF1co lane and hGORF1co lane) or cotransfected (hORF1co+hGORF1co lane) in HeLa cells. Western blot analysis of ORF1 protein species in HeLa cells under non-reducing (**left panel**) and reducing (**right panel**) conditions. Proteins are detected with human specific polyclonal antibodies. GAPDH used as loading control. A molecular weight ladder in kilodaltons (kDa) is presented on right side. Red asterisk indicates an ORF1p heterodimer. **C.** The same experiment as in B is performed in NIH-3T3 cells. Non-reducing western blot conditions are shown. Red asterisk denotes ORF1p heterodimers. **D.** Quantitation of western blot results from B and C using heterodimer/ORF1p monomer ratio. Statistical significance is indicated by dashed lines. An asterisk indicates statistically significant difference (T-test, p-value<0.05).

3.4.4: Heterodimer formation by the ORF1 proteins expressed from different mRNAs (heterodimerization *in trans)*

Our above-described approaches eliminated differing protein levels and subcellular localization as reasons contributing to the difference in the hORF1p self-interactions detected using the M2H assay in human and mouse cells (Figure 1). To identify the reason for this observed difference, we next determined which type of ORF1p interaction is responsible for the signal detected by the M2H system. Interactions between ORF1p homotrimers (*37*) as well as ORF1p self-interaction *in trans* could lead to transcriptional activation and expression of the reporter gene in the M2H system.

It is known that protein-protein interactions can be stabilized by disulfide bonds. It has been previously demonstrated that ORF1p trimerization can be stabilized by disulfide bonds *in vitro* (*37, 39*). Traditional reducing SDS-PAGE protein fractionation eliminates these bonds, subsequently leading to the detection of only ORF1p monomers. However, under non-reducing fractionation conditions, these covalent interactions are preserved. Western blot analysis following non-reducing SDS-PAGE of hORF1co and hGORF1co individually transfected in HeLa cells readily detected different species of ORF1p including monomers, homodimers, and homotrimers (Figure 5B). Homodimer size is specific to each protein because of the difference in the size of monomers (hGORF1co, 57 kDa, *versus* hORF1co, 40 kDa). If these proteins can form a heterodimer by interacting *in trans*, a unique band is expected to be detected in cells cotransfected with both expression plasmids (Figure 5A, middle lane). Indeed, coexpression of the tagged and untagged ORF1 proteins in HeLa and NIH-3T3 cells produced the predicted

ban of molecular weight consistent with a hORF1co/hGORF1co heterodimer (Figure 5B and C, asterisk). Figure 5D demonstrates that human cells support a more efficient ORF1p interaction *in trans* than do mouse cells, providing a plausible explanation for the difference in the ORF1p self-interaction detected by the M2H system (Figure 1B). ORF1p *trans* interaction is not unique to the protein produced by the codon-optimized expression plasmid, as coexpression of hORF1wt and hGORF1co proteins also leads to heterodimer formation in HeLa cells. The detection of this heterodimer using both human ORF1p- and T7-specific antibodies (hORF1wt contains a unique T7 tag) demonstrates the presence of the hORF1wt protein in this unique band (Supplemental Figure 5A and B). Together these data demonstrate that ORF1 proteins produced from different mRNAs can interact *in trans*.

The disulfide bonds that enable the above observations could have originated either in the cell or during cell lysis (*107-109*). To determine the origin of disulfide bonds between ORF1 proteins, total cell lysates were collected in the presence of N-ethylmaleimide (NEM), which prevents *de novo* disulfide bond formation. Western blot analysis following non-reducing SDS-PAGE separation of these samples demonstrated a significant reduction of observable ORF1p dimers and trimers (Supplemental Figure 6A), supporting that the majority of the disulfide bonds between ORF1p monomers occur during protein harvesting. Despite this, we contend that disulfide bond formation during lysis and harvest allows the capture of ORF1p self-interactions which occur within mammalian cells, as these bonds can only form between closely interacting proteins (*107*) (as also demonstrated by the following control experiments).

We wished to further test whether the disulfide bonds between the ORF1p monomers could be formed either during cell lysis or sample processing. We performed a control experiment by comparing HeLa cells either cotransfected with GAL4-fused or non-fused ORF1 expression vectors, or separately transfected with those vectors and subsequently mixed together (Supplemental Figure 6B). ORF1p heterodimerization may not be unique to the mammalian environment and the disulfide bonds may be formed during sample processing as opposed to during lysis of the cell. If either one of these possibilities is true, then the heterodimer band is expected to be detected by western blot analysis on the samples harvested from cells separately transfected with each expression vector and mixed prior to protein harvesting. Supplemental Figure 6C demonstrates that ORF1p heterodimerization is only detected when ORF1 proteins are coexpressed in mammalian cells, as no ORF1p heterodimerization was observed when cells separately expressing GAL4-fused or non-fused ORF1 proteins were combined prior to sample processing.

Furthermore, we investigated whether heterotrimer formation was possible outside the context of the intact cell as a function of time elapsed during sample processing. Protein samples harvested from cells transfected with either hORF1 or hGORF1 expression plasmid were mixed and incubated on ice in the presence or absence of NEM (Supplemental Figure 7A). Western blot analysis of these mixed samples demonstrated that the ORF1p heterodimers cannot form *in vitro* under the conditions tested, as no unique bands (predicted as a result of said heterodimer formation) were observed (Supplemental Figure 7B). Sonication of protein samples in the presence or

absence of NEM also had no effect on ORF1p heterodimer formation (Supplemental Figure 8A and B).

Combined, these results demonstrate that heterodimer formation by the ORF1 proteins produced by the M2H expression plasmids in mammalian cells is a definite source of the signal detected by the M2H system. However, it may not be the only source of the signal because we cannot rule out the possibility that interactions between homotrimers could potentially contribute to the signal. Importantly, ORF1p monomers can efficiently self-interact *in trans*. Quantitative analysis of heterodimerization efficiency between hGORF1coand hORF1co in HeLa and NIH-3T3 cells demonstrated that human cells supported ORF1p *trans* interaction more efficiently than did mouse cells (Figure 5D), providing a reasonable explanation for the difference in the ORF1p self-interaction between these cells as detected by the M2H system (Figure 1).

3.5: DISCUSSION

LINE-1 is the only currently active autonomous non-LTR retroelement in the human genome. It is responsible for the retrotransposition of itself and its parasites, the human SINE Alu and SVA elements (*7, 14, 62*). The L1 replication cycle begins with the transcription of a functional L1 locus and ends with an L1 copy integrated in a new genomic location. The necessary replication cycle steps include L1 mRNA gaining access to the cytoplasm, where translation of the two L1 encoded proteins (ORF1p and ORF2p) takes place, followed by assembly of the functional L1 RNPs composed of the L1 mRNA andL1-encoded proteins (*110*). These interactions exhibit a very strong *cis*-preference in mammalian cells (*36*). In the context of L1 retrotransposition, *cis*-preference is understood to be the propensity of the L1 translated proteins to associate with their encoding RNA (constituting the L1 RNP), as opposed to other RNAs (either cellular or generated by other L1 loci). Thus, according to this model, there is expected there is a preference for trimerization between ORF1 proteins made from the same mRNA. Through a yet unknown mechanism, the RNPs return to the nucleus and interact with genomic DNA to initiate integration through target primed reverse transcription (*30, 111*), a process which most likely depends on the many reported properties of the ORF1p (*38*).

Studies using *in vitro* systems have generated many important discoveries about human and mouse ORF1p self-interaction (*37, 39, 102*). The CCD motif of the mouse ORF1p was identified to be a requirement for trimerization (*29*). Despite the deletion of 7 out of 14 heptad repeats constituting the CCD, this observation was later confirmed for

the human ORF1p when its crystal structure was solved (*3*). A single cysteine present within the mouse ORF1p coiled-coiled domain was reported to be involved in the mouse ORF1p trimerization (*39*). The human L1 ORF1p contains four cysteine residues, whose involvement in the trimerization process is not yet established, as most of the coiled-coil domain of the crystallized human ORF1p was deleted.

Many (if not all) steps of the L1 replication cycle are downregulated by the host (*110*), most likely in an attempt by the host to mitigate the obvious potential for genomic instability that can arise from an overly enthusiastic retroelement (*15, 112*). Thus, understanding ORF1p self-interaction in the mammalian environment is important for identifying any potential cellular mechanisms controlling ORF1p assembly. We utilized a mammalian two-hybrid system that captures the ability of both human and mouse ORF1 proteins to self-interact *in trans* in cultured mammalian cells (Figure 1).

This system has many benefits for future L1 ORF1p studies, which are synergistically amplified when applied in combination with more traditional *in vitro* techniques. Among these advantages is the ease of the approach and detection of ORF1p self-interactions in the cellular environment. As the approach relies on transient transfections, the ORF1p self-interactions can be easily and rapidly tested in multiple cell types. The system does not require high protein expression levels, as both the codon-optimized and wild type mouse ORF1p produced a similar signal (Figure 1C). Furthermore, the effects of various treatments (particularly those affecting retrotransposition without altering L1 expression) on L1 ORF1p self-interaction could be conducted in different cell types. Finally, the system could also be used for detection and

mechanistic studies of ORF1p interactions with specific cellular proteins or for testing the effects of overexpression or depletion of specific cellular factors on L1 ORF1p interaction. Additionally, a positive readout in the M2H system mimics steps of the L1 replication cycle related to ORF1p biology.

As is the case with any experimental system, the M2H has its limitations. One of the main limitations of the system is that it cannot distinguish which ORF1p *trans*-interactions produce the positive signal, as *trans* interactions between monomers or homotrimers could contribute to the signal. Thus, the M2H assay cannot replace traditional *in vitro* approaches that clearly distinguish between these types of interactions. Furthermore, some wild type sequences, such as the wt human ORF1, may not produce any protein when subcloned into the M2H expression vectors, possibly due to the presence of functional splice sites (*93, 105, 106*) (Supplemental Figure 3A)

Characterization of ORF1p self-interactions using the M2H system resulted in a number of discoveries concerning ORF1p self-interaction. Although it has been previously reported that L1 proteins exhibit a strong *cis*-preference (*36*) (presuming that L1 ORF1p trimers are formed by the proteins translated from the same mRNA), we detected robust levels of ORF1p interaction between ORF1 proteins, which were produced from separate ORF1 expression plasmids (Figure 5 and Supplemental Figure 5). While these results may initially seem to contradict previous observations, we think that they expand our understanding of ORF1p biology without arguing against *cis*-preference. While the significance of *cis*-preference in the L1 replication cycle is well understood, the effects of ORF1p *trans* interactions remain almost completely unexplored

in respect to their significance to the L1 amplification. Indeed, it has already been reported that ORF1p enhances Alu retrotransposition (*15*), which is in itself reliant upon a type of *trans* interaction.

ORF1p *trans* interaction may have important biological implications for the L1 replication cycle. There are many more L1 loci in the human and rodent genomes with non-functional ORF1p than those that are competent for retrotransposition (*6*). Additionally, some prematurely polyadenylated and spliced L1 mRNAs (*77, 93*) have the potential to produce functional and non-functional ORF1 proteins. Many of these defective loci are expressed in human cells, possibly providing a pool of non-functional ORF1p that may interact *in trans* with functional ORF1p most likely creating non-functional trimers with an as-yet unknown effect on retrotransposition of functional L1 elements.

The non-reducing western blot analysis of the ORF1p self-interaction demonstrated that disulfide bonds formed during cell lysis stabilize ORF1p interactions preexisting within mammalian cells (Figure 5 and Supplemental Figures 5-8), potentially providing a glimpse into the coordination of ORF1 monomers within the trimeric structure. It is known that these cysteine residues must be in extremely close proximity to one another for such bonds to form. Thus, the formation of these bonds is indicative of the very tight association that occurs between ORF1p monomers within the cellular context. Equally important is the observation that heterodimers with resultant disulfide bonds cannot be formed by the simple mixing of cellular lysates during harvest, demonstrating the importance of the cellular context to heterodimer formation.

While characterizing ORF1p self-interaction in the M2H system, we also observed that a substantial proportion of hORF1p and mORF1p localize to the nucleus in both HeLa and NIH-3T3 cells (Figures 3, 4 and Supplemental Figures 3B, 3C, and 4). This pattern of subcellular localization was detected with GAL4-tagged ORF1 proteins expressed from codon-optimized vectors, and with the ORF1p generated from both wild type and codon-optimized ORF1 plasmids, and most importantly from wild type full-length L1 expression plasmids (Figure 4A and 4B). The dose curve experiment (Supplemental Figure 4D), along with the comparisons of ORF1p expression between codon-optimized and wild type expression plasmids (Figure 4), demonstrate that the predominantly nuclear localization of ORF1p is not concentration dependent within the parameters of our experiments. The time course experiment confirmed that the abundant nuclear presence of ORF1p is not time dependent (Supplemental Figures 4C and 4D). These findings are consistent with reports of nuclear detection of the endogenously expressed ORF1p in human tumor samples (*54, 55*) and the recently reported ORF1p interaction profile with nuclear and cytoplasmic proteins (*48*).

There are several technical differences between the previous studies detecting L1 ORF1p predominantly in the cytoplasm and those detecting it in both cellular compartments. Several previous publications detected L1 ORF1p in the cytoplasm of cultured cells using IHC (*54, 55, 103*). The experimental approach reported here relies on western blot analysis, which has very different sensitivity limits compared to IHC when detecting a potentially diffuse signal. We also utilized untagged ORF1p, cognizant of the fact that protein tags are notorious for their ability to alter mammalian protein function and/or subcellular localization (*113*). Consistent with these observations, fusion of the

VP16 tag to the N-terminus of human or mouse ORF1p significantly increased their cytoplasmic localization, despite the fact that a known NLS sequence is contained within the N-terminal VP16 fusion domain (Figures 1 and 3). This change in the subcellular localization is accompanied by reduced steady-state levels of the human and mouse VP16 ORF1 fusion proteins (Figure 3). While it was previously noted that some N-terminal fusions of ORF1p are not compatible with L1 retrotransposition (*103*), not every N-terminal fusion affects L1 ORF1p localization (Figure 3), suggesting that the altered subcellular localization is tag specific.

One of the other technical aspects of the two approaches (western blot and IHC) to consider is the sensitivity and specificity of the antibody. We cannot rule out the possibility that our antibodies detect nuclear and cytoplasmic forms of ORF1p, whether fused or unfused, with different efficiencies due to potential post-translational modifications. The same consideration applies to the IHC approach, where natively folded proteins are expected to be detected with antibodies recognizing denatured proteins. Additionally, ORF1p epitopes may also be obscured by protein-protein interactions, which are preserved when using IHC. This is especially concerning given that the L1 RNP (including all cellular factors) is largely poorly understood, with many potential ORF1p interacting partners having been recently identified (*48*). It is quite likely that ORF1p epitopes in the functional RNP may be obscured by attendant proteins preventing detection by the antibody. Our observations provide important experimental evidence that studies of tagged L1 proteins should be thoroughly validated, and comparisons of results obtained using different techniques should be interpreted with the consideration of the limitations associated with each approach.

Since the exact mechanism responsible for nuclear L1 ORF1p localization remains unknown, our findings have important implications for the L1 replication cycle. It is currently not known how the L1 RNP gains access to the nucleus. It was previously proposed that ORF2p contains an NLS, which may facilitate nuclear entry of the L1 RNP (*73*). Our data demonstrate that ORF1p may also contribute to L1 RNP entry into the nucleus, especially considering that it is likely that many more ORF1p than ORF2p molecules reside within an L1 RNP (*3, 37, 41*). This role of L1 ORF1p may be extended to other L1-dependent retroelements that parasitize the L1 mobilization machinery. ORF1p is known to enhance Alu retrotransposition (*15*), and our data open the possibility that this improvement may be due to the ORF1p-assisted entry of Alu RNPs into the nucleus.

In summary, we report an assay that provides a relatively easy analysis of L1 ORF1p self-interaction in mammalian cells. Our study demonstrates that transiently expressed mouse and human ORF1 protein is readily detected in the nuclear fractions of mammalian cells of human and mouse origin. This finding suggests that the ORF1p may be aiding in L1 and Alu RNP access to the nucleus. We also show that L1 ORF1 protein can interact *in trans*, raising the possibility of L1 retrotransposition interference by L1 loci expressing non-functional ORF1 proteins.

CHAPTER 4: TRUNCATED ORF1 PROTEINS CAN SUPPRESS LINE-1 RETROTRANSPOSITION IN *TRANS*

Published as: Mark Sokolowski, May Chynces, Dawn deHaro, Claiborne M. Christian and Victoria P. Belancio. Truncated ORF1 proteins can suppress LINE-1 retrotransposition in trans. *Nucleic Acids Research*. 2017;9(45):5294-5308. doi: 10.1093/nar/gkx211.

4.1: ABSTRACT

Long Interspersed Element 1 (L1) is an autonomous non-LTR retroelement that is active in mammalian genomes. Although retrotranspositionally incompetent and functional L1 loci are present in the same genomes, it remains unknown whether nonfunctional L1s have any *trans* effect on mobilization of active elements. Using bioinformatic analysis, we identified over a thousand of human L1 loci containing at least one stop codon in their ORF1 sequence. RNAseq analysis confirmed that many of these loci are expressed. We demonstrate that introduction of equivalent stop codons in the full-length human L1 sequence leads to the expression of truncated ORF1 proteins. When supplied in *trans* some truncated human ORF1 proteins suppress human L1 retrotransposition. This effect requires the N-terminus and coiled-coil domain as mutations within the ORF1p coiled-coil domain abolish the suppressive effect of truncated proteins on L1 retrotransposition. We demonstrate that the expression levels and length of truncated ORF1 proteins influence their ability to suppress L1 retrotransposition. Taken together these findings suggest that L1 retrotransposition may

be influenced by coexpression of defective L1 loci and that these L1 loci may reduce accumulation of *de novo* L1 integration events.

4.2: INTRODUCTION

Long INterspersed Element 1 (LINE-1, L1) is an autonomous, non-long terminal repeat retrotransposon that has contributed to the structural variability of mammalian genomes (*11*). L1 has a 5' untranslated region (UTR) followed by an Open Reading Frame 1 (ORF1), an inter-ORF region, an Open Reading Frame 2 (ORF2), and a 3' UTR with a polyA site and an associated polyA tail (*19, 20*). L1 transcription generates full-length mRNAs that produce two proteins, ORF1p and ORF2p (*21*). The ORF1p and ORF2p interact with their parental L1 mRNA in *cis* to form a ribonucleoprotein particle (RNP) (*7, 29*). All three RNP components are required for successful L1 retrotransposition (*7, 29, 110*). Strong *cis* preference of L1 proteins for their mRNA is beneficial for L1 retrotransposition and genome stability (*5, 36*). This *cis* preference minimizes the possibility for functional L1s to rescue retrotranspositionally incompetent L1s (*5, 36*), which are much more abundant than the active L1 loci (*5, 36*). It also minimizes nonspecific mobilization of cellular mRNAs (*5, 36*). However, L1-generated proteins do operate in *trans* when they mobilize non-autonomous human retroelements such as Alu and SVA (*14, 15, 62*). Although ORF2p alone is sufficient to mobilize Alu, ectopic L1 ORF1p expression enhances Alu mobilization through an unknown mechanism (*15*). Whether there is any *trans* effect of ORF1p expression on retrotransposition of human or mouse L1 elements remains unknown.

ORF1p has four recognized domains: a N-terminal domain, which contains two highly conserved phosphorylation sites critical for retrotransposition (*40*), a coiled-coil domain (C-C) containing a leucine zipper motif, an RNA recognition motif (RRM), and a C-terminal domain (*38*). These domains are present in both mouse and human ORF1 proteins (*3, 37, 39, 40*). Mouse and human ORF1 proteins function as homotrimers, which are formed through the coiled-coil domain (*3, 37, 39, 41-43*). The ORF1p also binds to RNA and has nucleic acid chaperone activity (*5, 29, 36, 39, 102*). ORF1p is generated in excess of the ORF2p (*3, 37, 41, 49*). Although ORF1p expression has been used as a correlative measure of L1 activity (*54, 55, 114, 115*), it remains unknown whether the amount of endogenous ORF1p adequately reflects L1 retrotransposition *in vivo*. Existing experimental data demonstrated that codon optimization of mouse and human L1 elements resulted in a substantial boost in L1 mRNA and protein expression of both L1 elements. This increase led to a dramatic increase in retrotransposition of the mouse L1 (*84*), but only a modest increase in the human L1 mobilization (*1, 78, 116*). Recently published data support that the relative ratio of L1 proteins may influence retrotransposition (*49*), with L1 RNPs containing few ORF1p being more efficient at integration relative to those containing abundant ORF1p trimers. It is also not known whether all generated ORF1p molecules retain strict *cis* preference for their parental L1 mRNA. The *trans* effect of the ORF1p on Alu retrotransposition (*15, 62*) and its requirement for SVA and host mRNA mobilization (*14*) demonstrates that some ORF1p is available to act in *trans* of their parental L1 mRNA.

L1 ORF1 protein homotrimerization has been observed *in vitro* (*3, 37, 39, 41, 42*). An interaction between mouse L1Tf and L1A ORF1 proteins has been observed using a

yeast two hybrid system (*43*). It has been proposed that ORF1p molecules generated from the same parental mRNA associate to form ORF1p homotrimers that are involved in retrotransposition in mammalian cells *(5)*. We have previously reported that ORF1p generated from different expression plasmids can form heterocomplexes in mammalian cells (*9*). This ability to heterotrimerize is species-specific, i.e. human L1 ORF1 proteins can heterodimerize with human (but not mouse) ORF1p in either mouse or human cells (*9*). This finding suggested the possibility that in mammalian cells, ORF1p produced from different L1 loci may have a *trans* effect on L1 retrotransposition in a manner similar to the effect of ORF1p on Alu mobilization (*15, 62*). The ability of defective L1 loci to *trans-complement* each other to restore efficient L1 retrotransposition was previously tested and was ruled out (*30, 36*). However, the possibility of the full-length ORF1p or truncated ORF1 proteins to affect retrotransposition of active L1s has not been investigated. The human genome contains thousands of full-length L1 loci (*10*), many of which have acquired premature stop codons within their ORF2 sequence (*10, 24-27, 66, 117, 118*). L1 mRNAs uniquely mapping to some of these retrotranspositionally-incompetent L1 loci were recovered from human cell lines (*117, 118*) suggesting the potential for existence of full-length L1 loci containing stop codons in their ORF1 sequence and capable of expressing truncated ORF1p.

RNAseq analysis of authentic endogenous L1 mRNAs from three human cell lines confirmed that L1 loci containing stop codons in their ORF1 sequence are expressed (*8*). Our data demonstrate that transient or stable expression of the full-length or truncated human ORF1 proteins suppresses human L1 retrotransposition in human and mouse cells. This suppressive effect is species-specific, as expression of the full-length or

truncated mouse ORF1p has no effect on human L1 mobilization. The dominant negative effect of the truncated human ORF1p on human L1 mobilization requires the N-terminus and an intact coiled-coil domain of the ORF1p. We demonstrate that mutagenesis of key leucine residues in the coiled-coil domain of the truncated human ORF1 protein abolishes its ability to form heterocomplexes and suppress L1 retrotransposition in *trans*. Confirmation of endogenous mRNA expression from L1 loci containing stop codons in their ORF1 sequence combined with our observations that stably expressed defective L1 elements suppress retrotransposition of transiently transfected active L1 in HeLa cells suggest that L1 loci expressing truncated ORF1 proteins may have a suppressive effect on retrotransposition of endogenously expressed functional L1 loci. These findings suggest that the unique spectrum of expressed non-functional L1 loci may differentially influence the efficiency of retrotransposition of functional L1s.

4.3: MATERIALS AND METHODS

4.3.1: Cells

HeLa (ATCC CCL2) and NIH-3T3 (ATCC CRL 1658) cells were maintained as previously described (*77*).

4.3.2: Plasmids

hL1wt is a plasmid expressing a full-length human wild-type L1.3 element (L1.3 plasmid as reported in (*36*)). L1Neo (pJM101/L1.3) (*7, 119*) is a plasmid designed to express a full-length human wild-type L1 element tagged with a neomycin resistance cassette, which enables detection of retrotransposition events upon transient transfection in mammalian cells. 119Stop (pJM108) (*7*) plasmid is designed to express a full-length untagged human L1.3 element containing a stop codon at the amino acid position 119 in ORF1. hORF1 (pBudORF1opt) (*15*) is a plasmid containing codon-optimized human L1 ORF1 sequence designed to express a full-length ORF1p. mORF1 (pBudORF1$_{syn}$) (*84*) is a plasmid containing codon-optimized mouse L1 ORF1 sequence designed to express a full-length mORF1p. mL1wt is a plasmid designed to express a full-length mouse wild-type L1spa element tagged with a neomycin resistance cassette, which enables detection of retrotransposition events upon transient transfection in mammalian cells (*82*). pBud (Invitrogen) was used as the empty plasmid control. pCEP (Invitrogen) was used as the empty control for Figures 7B, 7C, 8B and 8C. “pIRES” is pIRES2-GFP expression plasmid that was used in the toxicity assays (Supplementary Figures S2) and as a

transfection efficiency control (Figure 8C and Supplementary Figure S16B) (*2*). L1 “H1 H2” (L1PA1 Neo) (*78*), L1 “H1 M2” (pBS-L1-1H,2M*mneo*) (*78*), L1 “M1 H2” (pBS-L1-1M,2H*mneo*) (*78*) and L1 “M1 M2” (pBS syn mL1*mneo*) (*78*) are plasmids containing human and/or mouse codon-optimized ORF1 and ORF2 sequences. GAL4-fused ORF1 plasmid is designed to express a full-length human ORF1p containing a N-terminal GAL4 tag (*9*).

The DNA sequence of the 132M construct was synthesized (GenScript) to express a protein in which the leucine residues at positions 55, 73, 87, 90, 93, 100, 107 and 114 of human ORF1p were replaced with proline residues (CTG to CCC changes). The DNA sequence was subcloned into the pBud plasmid using HindIII and BamHI restriction endonucleases.

The DNA sequence of the mNhC-C construct was synthesized (GenScript) to contain codon-optimized sequences of the mouse ORF1 (corresponding to amino acids 1-46), human ORF1 (corresponding to amino acids 53-156), and a T7-tag. The DNA sequence of the hNmC-C construct was synthesized (GenScript) to contain codon-optimized sequence of the human ORF1 (corresponding to amino acids 1-52), mouse ORF1 (corresponding to amino acids 47-192), and a T7 tag. These sequences were subcloned into the pBud plasmid (Invitrogen) using HindIII and BamHI restriction endonucleases.

The previously reported consensus L1PA2 and L1PA3 ORF1 sequences (*81*) were used to synthesize (GenScript) full-length and truncated L1PA2 and L1PA3 ORF1 DNA sequence. The sequence AAG for lysine 14 was mutated to ACC to encode a threonine

residue in order to match the epitope sequence of our hORF1p antibodies (custom rabbit polyclonal, epitope: TGNSKTQSASPPPK) (*9*). The truncated ORF1 sequences corresponding to endogenous (e) L1 ORF1 e259 (L1PA2 subfamily, chromosome 13, position 37724090-37730119 of UCSC genome browser hg19 build), e207 (L1PA2 subfamily, chromosome 1, position 174812365-174818381 of UCSC genome browser hg19 build) and e127 (L1PA2 subfamily, chromosome 12, position 96709723-96715749 of UCSC genome browser hg19 build) were codon-optimized using Primo Optimum 3.4 (http://www.changbioscience.com/primo/primoo.html) and commercially synthesized (GenScript). The numbers in the names of these constructs (259, 207, and 127) correspond to the amino acid position of the stop codons present in endogenous L1 loci. Full-length and truncated L1PA2, L1PA3 as well as endogenous L1 sequences were subcloned into the pBud plasmid using HindIII and BamHI restriction endonucleases.

Unless specifically mentioned otherwise, truncated constructs were generated by PCR amplification of either wild-type (non-codon-optimized) or codon-optimized ORF1 DNA sequence. The resulting PCR products were subcloned into the pBud expression plasmid using HindIII and BamHI restriction endonucleases. See Supplementary Materials and Methods for primer sequences and generation of L1 expression plasmids containing stop codons.

4.3.3: Transfections

L1 retrotransposition assay: Adapted from (*7*), specific details for plasmid amounts, number of cells and transfection conditions for each figure are listed in the Supplementary Materials and Methods.

4.3.4: Western blot

4.3.4.1: Total protein extraction

The cellular lysates were processed as previously described (*9*). See the Supplementary Materials and Methods for details.

4.3.4.2: Toxicity Assay

Adapted from (*120*). See the Supplementary Materials and Methods for details.

4.3.4.3: Nuclear/Cytoplasm fractionation

The processing of nuclear and cytoplasmic fractions was performed as previously described (*9*). See the Supplementary Materials and Methods for details.

4.3.4.4: Western blot analysis

The western blot analysis was performed as previously described (*9, 121*). See the Supplementary Materials and Methods for details. The antibodies were diluted as listed below in a 3% milk (Bio Rad: 170-6404) in PBS-Tween: HRP-donkey anti-rabbit (Santa

Cruz; sc-2317), HRP-donkey anti-goat (Santa Cruz; sc-2020) or HRP-goat anti-mouse (Santa Cruz; sc-2031) at 1:5000 dilution. GAPDH antibodies (Santa Cruz: sc-25778, 1:5000 dilution) and Lamin A/C (Santa Cruz 7293, 1:1000 dilution) were used as a fractionation and equal loading controls. anti-hORF1p is a custom rabbit polyclonal antibody (epitope: TGNSKTQSASPPPK, dilution 1:5000) (*9*), anti-mORF1p is a custom goat-polyclonal antibody (epitope: YRTPNRLDQKRNSS, dilution 1:1000) (*9*), T7-tag antibody (Cell Signaling; D9E1X, 1:10000 dilution) and 1:10000 dilution of Flag tag (Sigma Monoclonal Anti-Flag M2: F3165) antibody.

4.3.5: Sucrose cushion of the cytoplasmic extract

See the Supplementary Materials and Methods for protocol details adapted from (*30, 78, 122, 123*)

4.3.6: Co-Immunoprecipitation

Adapted from (*122*). See the Supplementary Materials and Methods for protocol details.

4.3.7: Bioinformatic Analysis

L1Base (*6*) was used for analysis in Figure 6A and Supplementary Figure S10. L1Base (*6*) was utilized to identify full-length L1 loci containing stop codons in their ORF1 sequence. Human Full-Length >4,500nt LINE-1 Elements Ens38.36 was used as the database to identify the L1 loci containing no gaps or frameshifts in ORF1 and containing

one or more stop codons in ORF1, query start sites within 50 base pairs. This search identified 1474 L1 loci, which were then analyzed using the amino acid FASTA output for each of the ORF1 sequences to identify a methionine residue as the first amino acid of ORF1p. This approach identified 1244 L1 loci containing ORF1p with a start codon, which were further analyzed for the first stop codon position within the amino acid FASTA sequence. The frequency of stop codons at each ORF1 amino acid position was plotted (Figure 6A). The same approach was utilized for the mouse L1 loci using L1Base Mouse Full-Length >5knt LINE-1 Elements Ens38.35 (*6*). The first 10 L1 loci entries identified on each mouse chromosome (a total of 198 loci) were subjected to the first stop codon position analysis. The frequency of stop positions at each amino acid of the mouse ORF1 protein was plotted (Supplementary Figure S10). Data from (Deininger et. al., 2016) (*8*) was used for Supplementary Figure S13A and S13B. See Supplementary Materials and Methods for additional details.

4.4: RESULTS

4.4.1: Full-length human ORF1p suppresses human L1 retrotransposition in *trans*

We have previously reported that ORF1 proteins generated from different mRNA molecules can trimerize in a species-specific manner in mammalian cells (*9*). Based on these observations (*9*) and the report that the levels of ORF1p may influence L1 retrotransposition (*49*), we hypothesized that transient expression of the full-length human ORF1p may have a *trans* effect on retrotransposition of human L1. This hypothesis was tested using HeLa cells transiently cotransfected with plasmids expressing human neomycin-tagged L1 (L1Neo) and full-length human or mouse ORF1 proteins (hORF1p or mORF1p). The L1Neo expression plasmid contained wild-type L1 sequence. The ORF1 expression plasmids contained codon-optimized human or mouse L1 sequence. This approach demonstrated that when expressed in *trans* human, but not mouse, ORF1p suppressed human L1 retrotransposition (Figure 1A and Supplementary Figure S1). The same experiment was carried out in NIH 3T3 cells to test the possibility of a human-specific host factor being involved in the effect. Similar to human cells, human L1 retrotransposition was suppressed when this element was coexpressed with hORF1p, but not mORF1p, in NIH 3T3 cells (Figure 1B).

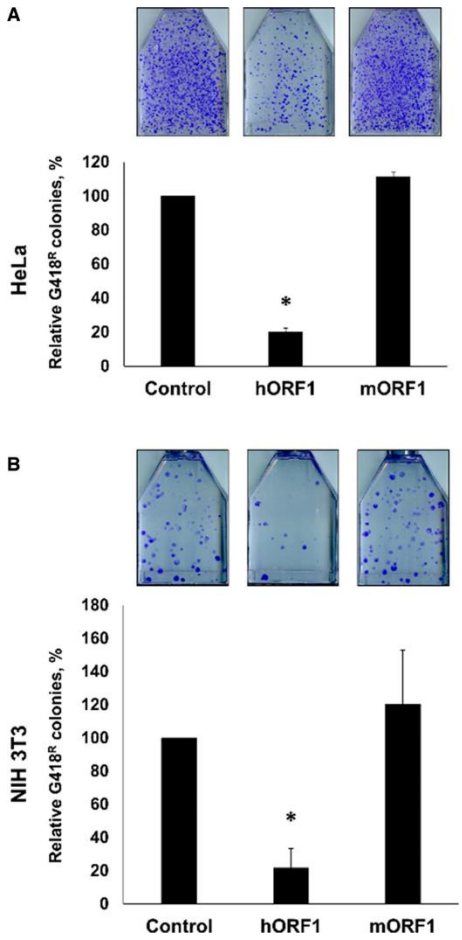

Figure 1. Full-length human ORF1p suppresses human L1 retrotransposition in *trans*. Results of L1 retrotransposition assay in HeLa **(A)** or NIH 3T3 **(B)** cells using neomycin-tagged, full-length, wild-type human L1 cotransfected with an empty (control), hORF1, or mORF1 expression plasmids. The number of G418 resistant colonies was normalized to the Control flask (L1 construct cotransfected with empty plasmid) for each independent experiment. Asterisks (*) denote statistical significance between listed constructs and the control (n=3, t-test, $p<0.05$).

To assess which of the L1-encoded proteins may be involved in the response to ORF1p *trans* effect on L1 retrotransposition we used previously reported chimeric L1 constructs containing codon-optimized ORF1 and ORF2 sequences of either human or mouse origin (Figure 2A) (*78*). Specifically, we used the H1H2 and M1M2 L1s, which contain only human or mouse sequences, respectively, the H1M2 L1, which contains human ORF1 and mouse ORF2 sequences, and the M1H2 L1 that contains mouse ORF1 and human ORF2 sequences. Transient transfections of HeLa cells with these plasmids and with either an empty plasmid (control), or plasmids expressing hORF1 or mORF1 proteins demonstrated that hORF1p suppressed mobilization of L1 constructs containing human ORF1 sequence (Figure 2B, black bars: H1H2 and H1M2). mORF1p had no negative *trans* effect on retrotransposition of human, mouse, or chimeric L1s (Figure 2B, dark gray bars).

The same experiment was also carried out in NIH 3T3 cells. Consistent with the results obtained in HeLa cells, cotransfection of hORF1 expression plasmid with the H1H2 or H1M2 L1 constructs containing human ORF1 sequence suppressed their retrotransposition (Figure 2C, black bars: H1H2 and H1M2). Expression of hORF1p did not have any effect on retrotransposition of M1M2 or M1H2 L1 elements. Cotransfection of the mORF1p expression plasmid with the M1M2 or H1M2 L1 constructs containing mouse ORF2 (mORF2) sequence resulted in a significant (5-6-fold) increase in their retrotransposition in NIH 3T3 cells relative to the control plasmid (Figure 2C, dark grey bars versus light gray bars).

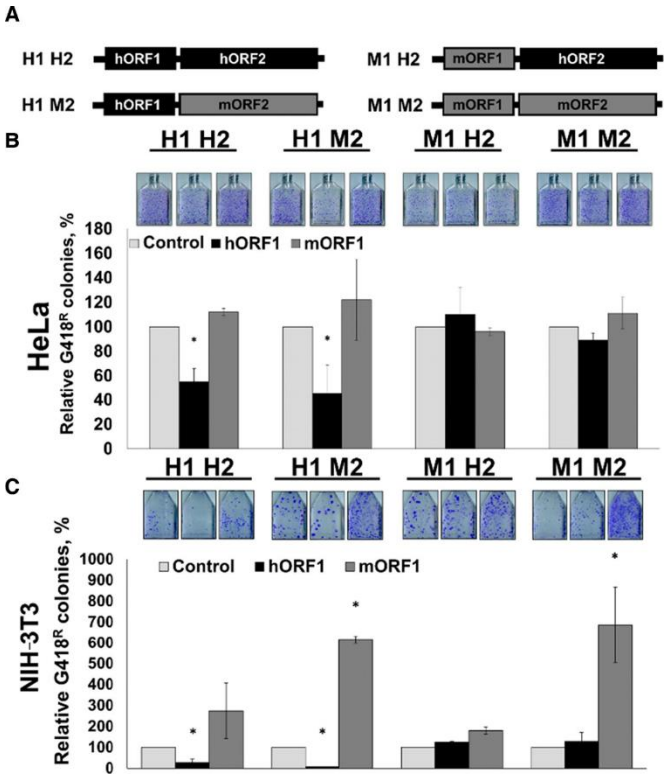

Figure 2. A full-length human ORF1p suppresses retrotransposition of chimeric L1s containing human ORF1 sequence. (A) Schematic of neomycin-tagged, full-length chimeric L1s containing human and/or mouse codon-optimized ORF1 and ORF2 sequences. Human ORF1 and human ORF2 are designated as H1 and H2, mouse ORF1 and mouse ORF2 are designated as M1 and M2, respectively. **(B)** Results of L1 retrotransposition assay in HeLa cells using plasmids expressing H1 H2, H1 M2, M1 H2 or M1 M2 L1s cotransfected with an empty (control), hORF1, or mORF1 expression plasmids. **(C)** Results of L1 retrotransposition assay performed as in (B) using NIH-3T3 cells. The number of G418 resistant colonies was normalized to the control flask (L1 construct cotransfected with an empty plasmid) for each independent experiment. Asterisks (*) denote statistical significance between listed constructs and the control (n=3, t-test, $p<0.05$).

Toxicity associated with human ORF1p expression, an increase in L1 toxicity in the presence of *trans* ORF1p, or ORF1p *trans* effect on retrotransposition through protein interactions with L1 or host proteins could explain the ORF1p-mediated decrease in L1 retrotransposition. ORF1p expression in *trans* did not have any adverse effect on cell viability when the full-length ORF1p was transiently expressed alone (Supplementary Figure S2A and S2B) or in combination with human L1 using a previously reported toxicity assay (*1, 2*) (Supplementary Figure S2C). In combination with the results shown in Figure 2, these findings support the involvement of the L1-encoded ORF1 protein in the *trans* effect of ORF1p on human L1 retrotransposition.

4.4.2: The ORF1p fragment containing the N-terminus and C-C domain is sufficient to suppress human L1 retrotransposition

To identify which region(s) of the hORF1p is responsible for the suppression of L1 retrotransposition and to test the possibility that truncated ORF1 proteins may also suppress retrotransposition, we generated 5 plasmids designed to produce C-terminally truncated human ORF1p, with the construct name indicating the amino acid position at the site of truncation (Figure 3A). Plasmids 54, 157, and 255 are designed to produce truncated human ORF1 proteins containing the N-terminus, N-terminus/coiled-coil, or N-terminus/coiled-coil/RNA binding domains, respectively. Plasmids 98 and 132 are designed to produce truncated human ORF1 proteins containing the N-terminus and 6 or 11 out of 14 heptads of the coiled-coil domain, respectively. A plasmid termed C-C is designed to express a T7-tagged coiled-coil domain of the human ORF1p because it is

involved in the ORF1p trimerization (*3, 37, 39, 41*). Western blot analysis using anti-ORF1p polyclonal antibodies recognizing an epitope in the N-terminus of the human ORF1p (*9*) or anti-T7 antibodies determined that other than the 54 construct, all plasmids expressed detectable levels of truncated ORF1 proteins in HeLa cells (Figure 3B). The 98 expression plasmid produced the lowest steady-state protein levels.

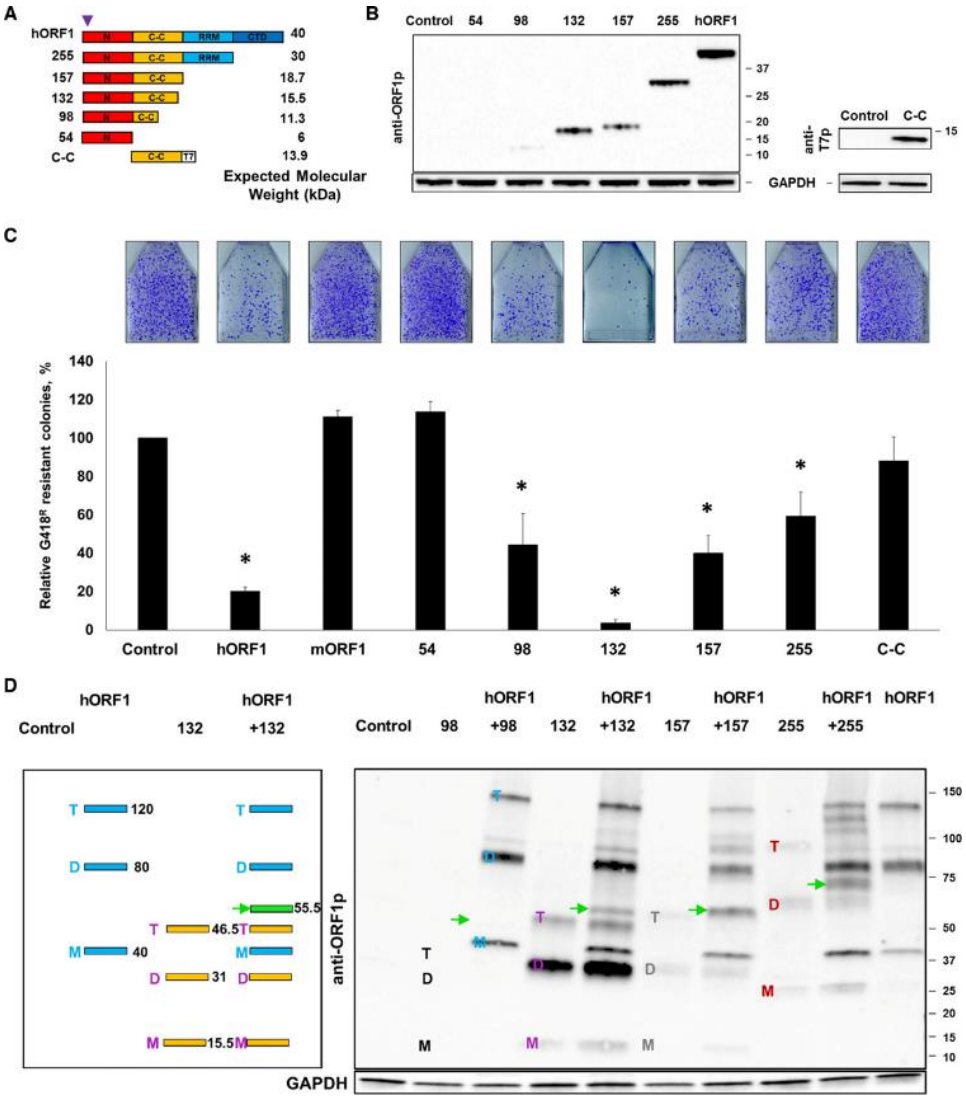
A
hORF1
255
157
132
98
54
C-C
N
C-C
RRM
CTE
T7
40
30
18.7
15.5
11.3
6
13.9
Expected Molecular Weight (kDa)
B
Control
54
98
132
157
255
hORF1
anti-ORF1p
37
25
20
15
10
anti-T7p
Control
C-C
15
GAPDH
C
Relative G418R resistant colonies, %
140
120
100
80
60
40
20
0
Control
hORF1
mORF1
54
98
132
157
255
C-C
D
hORF1
Control
132
hORF1
+132
T
120
D
80
55.5
T
46.5
M
40
D
31
M
15.5
Control
98
hORF1
+98
132
hORF1
+132
157
hORF1
+157
255
hORF1
+255
anti-ORF1p
T
D
M
150
100
75
50
37
25
20
15
10
GAPDH

Figure 3. Analysis of truncated ORF1 proteins expression and their potential to suppress L1 retrotransposition. (A) ORF1p domains are indicated as an N-terminal domain (N), a coiled-coil domain (C-C), an RNA recognition motif (RRM), and a C-terminal domain (CTD). The numbers listed on the left indicate the amino acid position of truncation. Expected molecular weights corresponding to each truncated protein are listed on the right. The approximate position of the epitope for the anti-ORF1p polyclonal antibodies is shown (purple triangle). The C-C construct has a T7 tag (T7). (**B) (left)** Western blot analysis of truncated ORF1 proteins transiently expressed in HeLa cells. Control lane indicates cells transfected with an empty plasmid. Numbers listed on the right correspond to molecular weight markers in kDa. Detection of GAPDH was used as a loading control. **(right)** Western blot analysis of C-C protein expression using anti-T7 tag polyclonal antibodies (anti-T7p). **(C)** Results of L1 retrotransposition assay in HeLa cells using a plasmid expressing a Neo-tagged, full-length human wild-type L1 cotransfected with an empty (control), hORF1, mORF1, 54, 98, 132, 157, 255 or C-C expression plasmids. The number of G418 resistant colonies was normalized to the control flask (L1 construct cotransfected with an empty plasmid) for each independent experiment. Asterisk (*) denotes statistical significance between listed constructs and the control (n=3, t-test, $p<0.05$). **(D) (left)** A schematic of the expected banding pattern resulting from the coexpression of the full-length (hORF1, blue) and 132 (132, yellow) ORF1 proteins when analyzed using non-reducing conditions. Blue and purple M, D, T letters and numbers (black) on the right correspond to monomers, dimers, trimers and their expected molecular weights. ORF1p 132 lane illustrates an appearance of a novel band with a unique molecular weight, if the coexpressed proteins form heterodimers (green band and arrow (55.5 kDa)). **(right)** Western blot analysis of proteins generated from codon-optimized 98, 132, 157, 255 and hORF1 constructs transfected individually or cotransfected in HeLa cells (non-reducing conditions, anti-ORF1p polyclonal antibodies (anti-ORF1p)). Green arrows indicate expected heterodimers between hORF1p and 98 (51.3 kDa), 132 (55.5 kDa), 157 (58.7 kDa) and 255 (70 kDa). M, D, T on

To assess the *trans* effect of these truncated human ORF1 proteins on human L1 retrotransposition, HeLa cells were transiently cotransfected with L1Neo and each of the above plasmids as described in Figure 1. This approach demonstrated that when expressed in *trans*, the 54 and C-C proteins did not suppress L1 retrotransposition (Figure 3C). Cotransfection of the 98, 132, 157, or 255 expression constructs with the human L1Neo significantly reduced L1 mobilization, with the 132 truncated protein being the most efficient in suppressing L1 mobilization in *trans*. The 132 construct reduced L1 retrotransposition to the lowest levels at 4% of the control, which surpassed the extent of inhibition observed with the full-length ORF1p expression plasmid (Figure 3C).

A potential mechanism that could explain this negative *trans* effect of the truncated ORF1 proteins on L1 retrotransposition is through a direct interaction between the truncated ORF1p and the full-length ORF1p, which would produce ORF1p trimers containing one or two defective monomers. To test whether *trans* interactions between the truncated and full-length ORF1 proteins exist, HeLa cells were transiently cotransfected with plasmids containing truncated ORF1 sequences and an expression plasmid containing full-length ORF1 sequence. Cotransfection of these plasmids allows for an assessment of the ability of the proteins they produce to heterodimerize using a previously reported western blot analysis carried out under non-reducing conditions (*9*). The premise is that the cysteine residues present in the C-C domain of the human ORF1 protein cross link interacting ORF1 proteins upon oxidation during protein harvest resulting in bands with unique molecular weights corresponding to these heterodimers (Figure 3D, a green band). Consistent with their ability to suppress L1 mobilization in

trans, the 132, 157, and 255 truncated ORF1 proteins formed heterodimers of expected molecular weights with the full-length ORF1p (Figure 3D, green arrows). Because the 98 construct expressed significantly less protein than the 132, 157, or 255 constructs when analyzed using reducing conditions for western blot analysis (Figure 3B), our anti-ORF1p antibodies were not sensitive enough to detect monomers, dimers, trimers, and heterodimers formed by this truncated protein under non-reducing conditions (Figure 3D). The C-C construct did not form heterodimers with the full-length ORF1p when coexpressed in HeLa cells, consistent with its inability to suppress L1 retrotransposition (Supplementary Figure S3).

Combined, these data demonstrate that the ORF1p fragment containing the N-terminus and coiled-coil domains of the human ORF1 protein is sufficient to suppress human L1 retrotransposition in *trans*. Our results also demonstrate that the inclusion of the RRM domain reduces the suppressive effect of truncated ORF1p on L1 mobilization. Also, our finding that the C-C domain alone does not suppress L1 retrotransposition in *trans*, suggests that the N-terminus may be important for this effect (Figure 3C).

4.4.3: Truncated chimeric ORF1 proteins composed of human and mouse sequences do not suppress human L1 retrotransposition

The fact that a truncated ORF1p containing only the N-terminal sequence of the ORF1 protein is not stable in HeLa cells prevented us from testing its sole effect on the suppression of L1 retrotransposition. To test the contribution of the N-terminal domain to the ORF1p mediated *trans* effect on retrotransposition by other approaches, we

generated constructs designed to express chimeric proteins hNmC-C and mNhC-C. The hNmC-C protein contains the N-terminal sequence of the human ORF1p and the coiled-coil domain of the mouse ORF1p. The mNhC-C protein contains the N-terminus of the mouse ORF1p and the coiled-coil domain of the human ORF1p (Supplementary Figure S4A). Both chimeric constructs contain a T7 tag to allow for their detection by western blot analysis. Western blot analysis of total cellular lysates collected from HeLa cells transiently transfected with the chimeric plasmids detected chimeric proteins of expected molecular weights (Supplementary Figure S4A). Western blot analysis using non-reducing conditions demonstrated that the mNhC-C protein did not heterodimerize with the full-length hORF1p when coexpressed in HeLa cells (Supplementary Figure S4B). These chimeric NC-C constructs had no effect on L1 retrotransposition when cotransfected with the Neo-tagged human L1 expression plasmid in HeLa cells (Supplementary Figure S4C). These findings, combined with the lack of the suppressive effect of the truncated C-C protein on L1 retrotransposition, support that the N-terminus as well as the coiled-coil domain of the human ORF1p are needed to reduce L1 retrotransposition in *trans*.

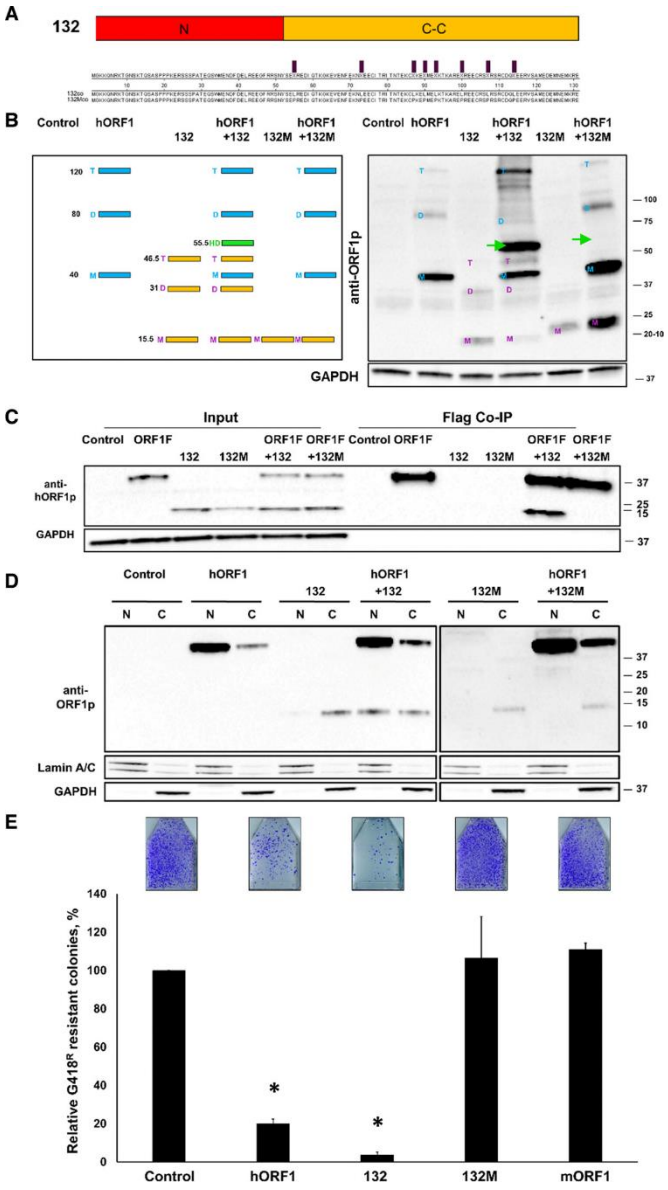
A
132
N
C-C
B
Control
hORF1
132
hORF1 +132
132M
hORF1 +132M
anti-ORF1p
GAPDH
C
Input
Flag Co-IP
Control
ORF1F
anti-hORF1p
GAPDH
D
N
C
anti-ORF1p
Lamin A/C
GAPDH
E
Relative G418R resistant colonies, %
Control
hORF1
132
132M
mORF1

Figure 4. Mutations in the coiled-coil domain of truncated ORF1 protein abolishes its suppressive effect on L1 retrotransposition. (A) A schematic of the 132 construct (N-terminus and coiled-coil domain) and alignment of functional (132) and mutant (132M) protein sequences. The positions of 8 leucine residues mutated to proline residues are indicated by purple bars. **(B) (left)** A schematic of the expected band pattern resulting from heterodimerization of the full-length ORF1p and truncated 132 ORF1p (green band). Numbers indicate molecular weights (kDa) for monomers (M), dimers (D), and trimers (T) and heterodimers (green band). **(right)** Western blot analysis using anti-ORF1p polyclonal antibodies (anti-ORF1p) of proteins generated from codon-optimized 132, hORF1, and 132M constructs transiently transfected in HeLa cells (non-reducing conditions). Green arrows indicate the expected position of the heterodimer (55.5 kDa). Numbers on the right of the image are molecular weight markers. GAPDH is used as a loading control. **(C)** Co-immunoprecipitation of the Flag-tagged full-length ORF1p (ORF1F) with the wild-type (132) or mutant (132M) truncated ORF1p. Western blot analysis was performed using anti-ORF1p polyclonal antibodies (anti-ORF1p). Control lane indicates cells transfected with an empty plasmid. GAPDH is used as a loading control. Positions of molecular markers are indicated on the right in kDa. **(D)** Western blot analysis (reducing conditions, human-specific ORF1p polyclonal antibodies) of nuclear (N) and cytoplasmic (C) fractions collected from HeLa cells expressing hORF1, 132co, or 132Mco proteins. GAPDH and Lamin A/C were used as loading controls. **(E)** Result of L1 retrotransposition assay using a Neo-tagged, full-length human wild-type L1 expression plasmid cotransfected with an empty (control), hORF1, 132, 132M or mORF1 expression plasmids. The number of G418 resistant colonies was normalized to the Control flask for each independent experiment. Asterisk (*) denotes statistical significance between listed constructs and the control ($n=3$, t-test, $p<0.05$).

4.4.4: Mutations within the C-C domain of the truncated human ORF1p abolish its ability to suppress human L1 retrotransposition

The confirmation that the N-terminal half of the human L1 ORF1p protein is sufficient to suppress L1 retrotransposition in *trans* prompted genetic manipulations to identify the requirements for this effect. To test the hypothesis that efficient heterodimerization through the C-C domain is required for the ORF1p suppressive effect on L1 retrotransposition, 8 leucine (L) amino acids present in the coiled-coil domain of the human ORF1p were mutated into proline (P) amino acids to generate the 132 mutant construct (132M) (Figure 4A). This approach was chosen based on the previously published work demonstrating that the L to P mutations within coiled-coil domains or leucine zipper motifs of other proteins abolished their interactions (*124-126*). Western blot analysis using anti-ORF1p polyclonal antibodies confirmed that the 132M protein was expressed in HeLa cells (Figure 4B, 132M lane in the right panel). Western blot analysis of cellular lysates harvested using non-reducing conditions demonstrated that the 132M protein was unable to form heterodimers with the hORF1p when both proteins were coexpressed in HeLa cells (Figure 4B, hORF1/132M lane in the right panel). The presence of the heterodimeric complexes formed by the wild-type 132 and full-length ORF1 proteins was readily detected using the same experimental conditions (Figure 4B, green arrows). To further confirm the loss of interaction between the 132M and full-length ORF1 proteins, we utilized a coimmunoprecipitation (co-IP) approach. We generated a C-terminally fused Flag-tagged ORF1 construct and confirmed that the protein generated from this construct heterodimerizes with the protein generated from a construct containing GAL4-fused ORF1 sequence in HeLa cells using non-reducing

western blot approach (Supplementary Figure S5). HeLa cells were cotransfected with this Flag-tagged ORF1p expression plasmid and either the wt or mutant 132 expression plasmids. The resulting protein lysates were subjected to a pull down using anti-Flag antibody. The input and pull-down protein fractions were analyzed by western blot analysis using ORF1-specific polyclonal antibodies (*9*). This approach demonstrated that that the wild-type 132 protein co-IPed with the full-length human ORF1p, while the 132M protein did not (Figure 4C).

To further characterize the 132p/ORF1p interactions we determined their subcellular localization. Western blot analysis using anti-ORF1p polyclonal antibodies demonstrated that the full-length ORF1p was observed in the nuclear fraction of HeLa cells transfected with the human ORF1 expression plasmid, consistent with our previous observations (*9*). The truncated, functional and mutant 132 proteins were detected predominantly in the cytoplasmic fraction of HeLa cells transfected with these respective expression plasmids (Figure 4D, lanes 132 and 132M). Cotransfection of the wilf-type 132 and full-length ORF1 plasmids in HeLa cells resulted in the detection of the wild-type 132 protein in the nuclear fraction (Figure 4D, compare lanes 132 and ORF1+132). In contrast, the 132M protein remained cytoplasmic in the presence of the full-length ORF1p (Figure 4D, compare lanes ORF1+132 and ORF1+132M). Consistent with its inability to heterodimerize with the full-length ORF1p as detected by western blot analysis, co-IP, and subcellular localization, the 132M protein had no effect on L1 mobilization when transiently coexpressed with the L1Neo in HeLa cells (Figure 4E).

4.4.5: Expression levels and combination of truncated ORF1 proteins influence their *trans* effect on human L1 retrotransposition

To determine whether expression levels alter the efficiency with which the truncated ORF1 proteins suppress L1 retrotransposition in *trans*, we generated C-terminally truncated 157, 132 and 98 ORF1p expression constructs shown in Figure 3A utilizing human non-codon-optimized wild-type ORF1 sequence (Figure 5A). Western blot analysis with anti-ORF1p polyclonal antibodies detected that as expected the steady-state levels of all proteins produced by the wild-type constructs were reduced compared to the constructs containing codon optimized L1 ORF1 sequences (Figure 5B and Supplementary Figure S6). Very low 98wt and 157wt protein expression resulted in the corresponding loss of their suppressive *trans* effect on L1 retrotransposition using 1:1 ratios of the transfected plasmids (Figure 5C). Despite its reduced expression, the 132wt construct still significantly suppressed L1 retrotransposition (Figure 5C, 132wt).

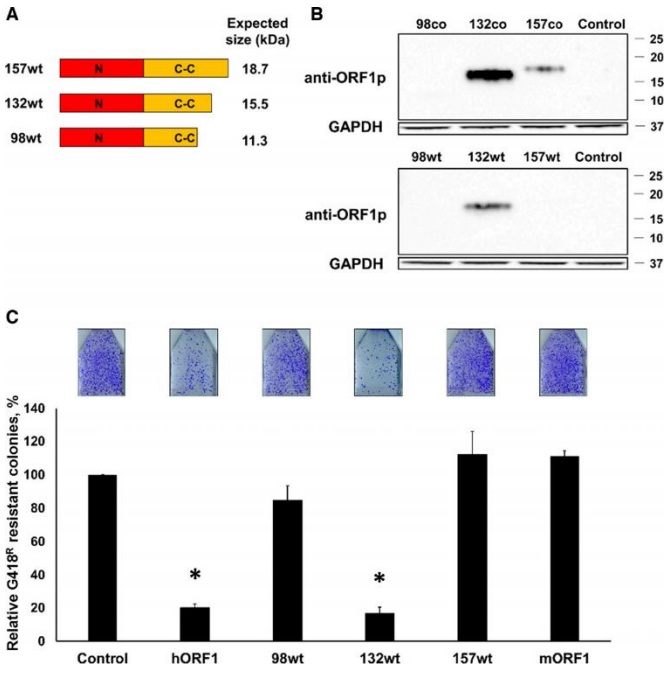

Figure 5. Expression and effect of truncated ORF1 proteins generated from plasmids containing wild-type L1 sequence on L1 retrotransposition. (A) A schematic of the C-terminally truncated ORF1 expression constructs containing wild-type L1 sequence (98wt, 132wt and 157wt). **(B)** Western blot analysis using ORF1-specific antibodies of truncated ORF1 proteins generated from codon-optimized (co) or wild-type (wt) 98, 132, 157 constructs transiently transfected in HeLa cells. Control lane indicates cells transfected with an empty plasmid. GAPDH is used as a loading control. Positions of molecular weight markers are indicated on the right in kDa. **(C)** Result of L1 retrotransposition assay using a Neo-tagged, full-length, human wild-type L1 expression plasmid cotransfected with an empty (control), hORF1, 98wt, 132wt, 157wt or mORF1 constructs. The number of G418 resistant colonies was normalized to the control flask for each independent experiment. Asterisk (*) denotes statistical significance between listed constructs and the control (n=3, t-test, $p<0.05$).

It has been previously reported that both functional and retrotranspositionally-incompetent L1 loci are coexpressed (*117, 118*). These published results suggest that the relative ratio of the functional and non-functional L1 loci as well as the make-up of truncated ORF1p expressed from non-functional L1 loci may vary among cells. To test the effect of coexpression of different amounts and forms of truncated ORF1 proteins on L1 retrotransposition, 0.2 or 0.4µg of the 98wt and 157wt plasmids were cotransfected individually or together with the Neo-tagged human L1 in HeLa cells (Supplementary Figure S7A). The coexpression of 98wt and 157wt truncated ORF1 proteins with the Neo-tagged human L1 suppressed L1 retrotransposition when compared to the individual effects of each truncated ORF1 protein under the same transfection conditions (98wt or 157wt) (Supplementary Figure S7B). The same result was observed when truncated ORF1 proteins produced from the plasmids containing codon-optimized sequences were coexpressed (98co and 157co). As with the wild-type proteins, the coexpression of 98co and 157co proteins resulted in a greater suppression of L1 retrotransposition than their individual ability to limit L1 mobilization under the same transfection conditions (Supplementary Figure S7B). These results support that a specific combination of truncated ORF1 proteins expressed from multiple L1 loci as well as the levels of their expression may differentially influence L1 retrotransposition.

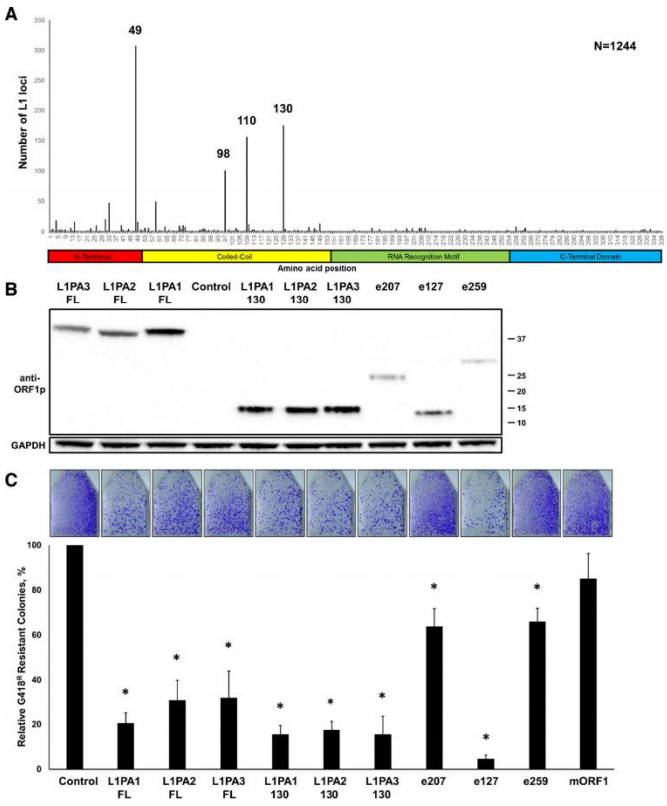

Figure 6. Bioinformatic analysis of stop codon positions in retrotranspositionally-incompetent human L1 loci. (A) Positions and frequencies of stop codons identified in the ORF1p sequence of 1244 full-length human L1 loci. The most frequent stop codon positions corresponding to amino acids 49, 98, 110 and 130 of the full-length human ORF1p are indicated. **(B)** Western blot analysis using anti-ORF1p Ab of full-length (FL) and truncated (130) ORF1 proteins generated from expression plasmids containing consensus sequences corresponding to L1PA1-3 subfamilies or sequences representing endogenously expressed L1s (e207, e127 or e259). Control lane indicates cells transfected with an empty plasmid. Positions of molecular markers are indicated on the right in kDa. **(C)** Results of L1 retrotransposition assay using a Neo-tagged, full-length human wild-type L1 expression plasmid cotransfected with an empty (Control), L1PA1-3 full-length (FL) or truncated (130, e207, e127 or e259) human ORF1 expression plasmids or a mORF1 expression plasmid. The number of G418 resistant colonies was normalized to the Control flask for each independent experiment. Asterisk (*) denotes statistical significance between listed constructs and the Control (n=3, t-test, $p<0.05$).

4.4.6: Bioinformatic analysis of genomic L1 loci identifies species-specific distribution of stop codon positions

It has been previously reported that retrotranspositionally incompetent L1 loci are expressed (*117, 118*). However, it has not been determined how many full-length L1 loci containing stop-codons in their ORF1 sequence are present in the human genome. We analyzed the human genome to identify full-length L1 loci harboring stop codons within their ORF1 sequence that would have the potential to generate truncated ORF1 protein. Using L1Base (*6*), we identified 1244 human L1 loci containing ORF1 with one or more stop codons. These 1244 L1 loci were further analyzed to identify the position of their first stop codon in the ORF1 sequence. This analysis determined that the majority of identified human L1 loci (59%) contained their first stop codon at amino acid positions 49, 98, 110 or 130 (Figure 6A). The distribution of codons that can give rise to stop codons within the human ORF1 sequence via a single point mutation is shown in the Supplementary Figure S8. Out of the 1244 total L1 loci containing at least one stop codon in their ORF1 sequence 375 L1 loci harbored a stop codon within the sequence corresponding to the 99 to 132 amino acid region of the full-length ORF1p. Based on our experimental data, ORF1 proteins truncated within this region could be the most efficient in suppressing L1 retrotransposition in HeLa cells (Figures 3D and 5C). Bioinformatic analysis of these 375 loci determined that over 75% of them belonged to the three youngest L1 subfamilies (L1PA1-3). Specifically, 8.5% of loci belonged to the L1 HS subfamily, 35.7% belonged to the L1PA2 subfamily, and 32.8% belonged to the L1PA3 subfamily (Figure 6A). Using the UCSC browser, we also identified that in four out of randomly chosen 50 full-length L1 loci, the presence of the stop codon in the ORF1

sequence was polymorphic (Supplementary Figure S9). The estimated allele frequencies of the L1 variant containing the stop codon ranged from 0.1 to 0.58 (Supplementary Figure S9).

The above described approach was also utilized to identify mouse L1 loci fitting the same criteria. We chose the first 10 L1 loci entries identified on each mouse chromosome, which resulted in the total of 198 loci for further analysis of the position of the first stop codon in their ORF1 sequence. Among mouse L1 loci, the most common position for the first stop codon was at amino acid 251 (Supplementary Figure S10). Collectively, our bioinformatic analysis of the mouse and human L1 loci demonstrates that many full-length human and mouse L1 loci present in their respective genomes contain premature stop codons at specific dominant positions within their ORF1 sequence and that these prevalent positions differ between mouse and human elements. All of these full-length L1 loci are retrotranspositionally-incompetent, yet they have the potential to be transcribed and translated to generate truncated ORF1 proteins with a possible suppressive effect on retrotransposition of functional L1 loci.

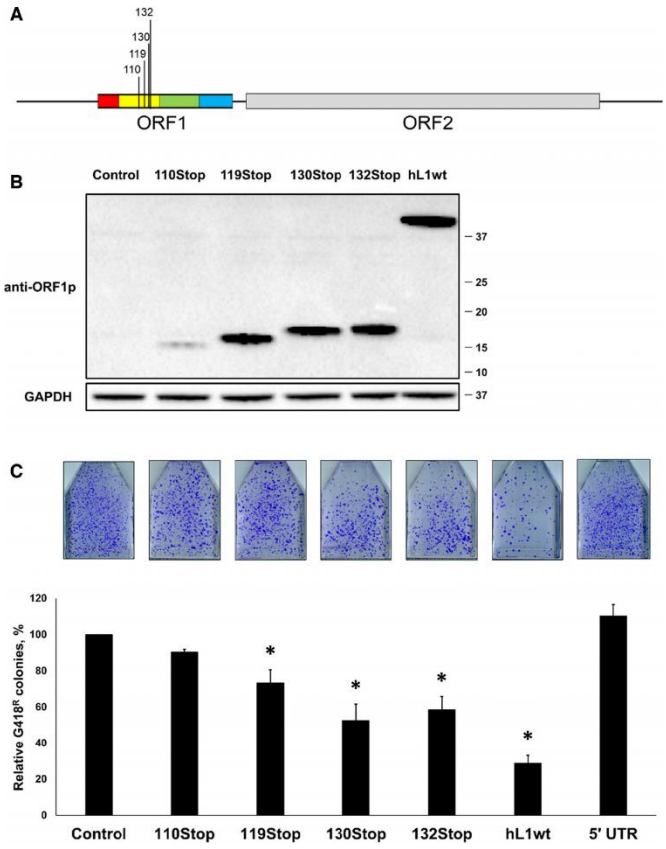

Figure 7. Human L1 retrotransposition is suppressed by full-length human L1s containing ORF1p stop codons. (A) A schematic of the L1 constructs generated to contain stop codons corresponding to amino acid positions 110, 119, 130 and 132 within the human L1 ORF1. **(B)** Western blot analysis (anti-ORF1p polyclonal antibodies) of ORF1 proteins generated from plasmids expressing full-length wt L1 (hL1wt) or full-length L1s containing stop codons (110Stop, 119Stop, 130Stop, 132Stop) in HeLa cells. A transfection with an empty plasmid was used as control (Control). Detection of GAPDH was used as a loading control. Molecular weight markers are shown on the right in kDa. **(C)** Result of L1 retrotransposition assay in HeLa cells using a Neo-tagged full-length human wt L1 expression plasmid co-transfected with an empty (control), hORF1, mORF1, 110Stop, 119Stop, 130Stop or 132Stop L1 expression plasmids. The number of G418 resistant colonies was normalized to the number of colonies determined for the Control flask for each independent experiment. Asterisk (*) denotes statistical significance between listed constructs and the control (n=3, t-test, $p<0.05$).

In the above described experiments (Figures 3, 4, 5 and Supplementary Figure S7), we determined that various truncated human ORF1 proteins suppress human L1 retrotransposition. However, whether truncated mouse ORF1 proteins behave in a similar manner is not known. Based on the L1Base analysis of selected mouse L1 loci shown in the Supplementary Figure S10, we generated a construct designed to express a 250 amino acid long mouse ORF1 protein tagged with a T7 sequence (m251) (Supplementary Figure S11A). Western blot analysis using polyclonal anti-mORF1p (*9*) or anti-T7 antibodies detected truncated m251 protein expression in HeLa cells (Supplementary Figure S11B). To test its ability to *trans*-interact with the full-length mORF1p, the m251 expression plasmid was cotransfected with a full-length mORF1 plasmid in HeLa cells. Western blot analysis using non-reducing conditions and either anti-mORF1 (*9*) or anti-T7 tag polyclonal antibodies detected a unique band that was consistent with the expected molecular weight of a m251/ORF1 protein heterodimer (Supplementary Figure S11C, green arrows). Transient cotransfection of HeLa cells with the neomycin-tagged mouse L1 and the m251 expression constructs demonstrated that when expressed in *trans*, the m251 protein significantly suppressed mouse L1 retrotransposition (Supplementary Figure S11D, m251). As expected, coexpression of the m251 protein with the Neo-tagged human L1 did not affect human L1 retrotransposition (Supplementary Figure S12). Combined these data demonstrate that, when coexpressed in *trans*, specific truncated mouse and human ORF1 proteins have a similar suppressive "dominant-negative" effect on the retrotransposition of their respective full-length L1 elements.

4.4.7: L1 loci containing one or more stop codons within the ORF1 sequence are expressed in human cell lines

In order to determine whether any of the L1 loci containing stop codons in their ORF1 sequence identified in the human genome are expressed in human cells, we utilized an RNAseq approach that identifies individual endogenous L1 loci expressed in HeLa and HEK 293 cells (*8*).We determined the subfamily of the top 50 endogenous L1 loci identified to be expressed in HEK 293 or HeLa cells (*8*) analyzed them for the presence of stop codons in their ORF1 sequence. This analysis determined that the majority (49/50) of the expressed L1 loci identified in HEK 293 cells belonged to three youngest L1 subfamilies (the L1 HS 16/50, L1PA2 23/50 and L1PA3 10/50) (Supplementary Figure S13A). The endogenous L1 loci expressed in HeLa cells belonged to PA1-PA6 subfamilies (L1HS 1/50, L1PA2 4/50, L1PA3 9/50, L1PA4 16/50, L1PA5 15/50 and L1PA6 5/50) (Supplementary Figure S13B). This analysis determined that the majority of endogenously expressed L1 loci containing stop codons in the amino acids 99 through 132 region of the ORF1p belonged to the L1PA2 or L1PA3 subfamilies (Supplementary Tables 1 and 2). In agreement with RNAseq results identifying endogenous L1 mRNA expression, western blot analysis using anti-ORF1p polyclonal antibodies detected a band consistent with the expected size of the full-length ORF1 protein in HEK 293 cells (Supplementary Figure S14). No bands consistent with truncated ORF1 proteins were detected. Analysis of protein sequence corresponding to the top 50 L1 loci identified by RNAseq (*8*) determined that 66% of these loci code for proteins that have 1 or more subsititutions within the epitope recognized by our Abs. Most of the L1 loci containing no stop codons within the ORF1 sequence had a perfect match with the epitope. These

findings demonstrate that retrotranspositionally competent L1 loci and L1 loci containing stop codons in their ORF1 sequence are coexpressed in mammalian cells.

4.4.8: ORF1p generated from L1PA2 and L1PA3 suppress L1 retrotransposition in *trans*

The RNAseq analysis determined that many endogenously expressed L1 loci harboring stop codons in their ORF1 sequence belong to the L1PA2 and L1PA3 subfamilies (Supplementary Tables 1 and 2). To determine whether truncated ORF1 proteins generated by the L1PA2 and L1PA3 subfamilies can suppress L1 retrotransposition as effectively as the modern L1 ORF1, we generated plasmids containing commercially synthesized full-length or truncated ORF1 consensus sequences representing the L1PA2 and L1PA3 elements (*81*). We also generated three additional plasmids containing ORF1 sequences corresponding to specific endogenous L1PA2 loci identified in our analysis of the RNAseq expression data from HEK 293 cells (*8*). These plasmids referred to as endogenous (e) e207, e259, and e127 constructs are expected to produce truncated ORF1 proteins of 206, 258, or 126 amino acid long, respectively (Supplementary Table 1). Western blot analysis using anti-ORF1p polyclonal antibodies determined that all plasmids expressed detectable levels of ORF1 proteins of the expected molecular weights when transiently transfected in HeLa cells (Figure 6B). Western blot analysis of cellular lysates harvested using non-reducing conditions demonstrated that the full-length L1PA2 and L1PA3 ORF1p readily formed heterodimers with the GAL4-fused L1PA1 ORF1p when coexpressed in HeLa cells (Supplementary Figure S15A, green arrows). The truncated L1PA1, L1PA2, L1PA3 ORF1p also formed heterodimers with

the full-length L1PA1 ORF1p when coexpressed in HeLa cells (Supplementary Figure S15B, green arrows). Consistent with these findings, both the full-length and truncated (130 aa) PA1, 2, and 3 proteins efficiently suppressed modern L1 mobilization when transiently coexpressed with the L1Neo in HeLa cells (Figure 6C).

The e259, e207 and e127 proteins produced by their respective plasmids were expressed at lower levels than the L1PA1-3 130 proteins (Figure 6B). They formed heterodimers with the full-length modern ORF1p when coexpressed in HeLa cells (Supplementary Figure S15C, green arrows) and suppressed L1 mobilization with various efficiencies (Figure 6C). The very efficient suppression of the L1.3 retrotransposition by the e127 protein compared to the statistically significant, but inefficient, suppression of L1 retrotransposition by the e207 and e259 proteins is consistent with our data shown in Figures 3B, 3C, 5B and 5C. These results show that truncated ORF1 proteins with lower expression levels and longer truncated ORF1 proteins are less efficient at suppressing L1 retrotransposition than the ORF1p truncated around the amino acid position 130. These data demonstrate that the full-length and truncated ORF1 proteins representing L1PA2 and L1PA3 subfamilies are able to efficiently interact with the modern L1 ORF1p and suppress L1 retrotransposition in *trans*.

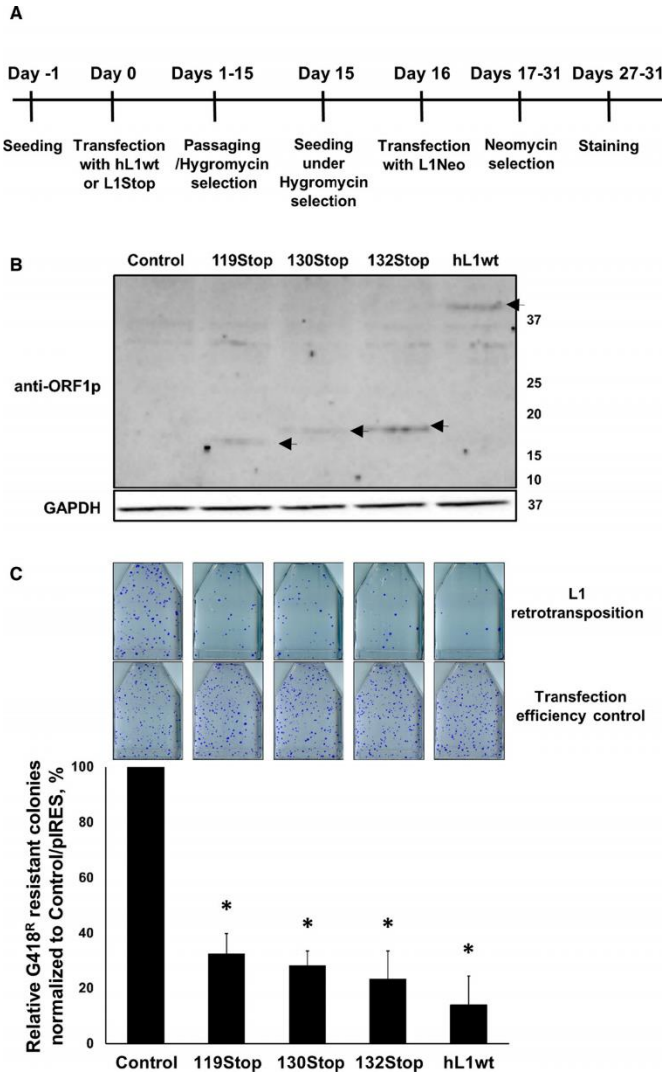

Figure 8. L1 retrotransposition is suppressed in HeLa cells constitutively expressing a full-length human L1 or full-length human L1s containing stop codons in their ORF1 sequence. (A) Schematic of experimental design. **(B)** Western blot analysis (anti-ORF1p polyclonal antibodies) of ORF1 proteins in engineered HeLa cells harboring full-length L1 (hL1wt), full-length L1s with stop codons (119Stop, 130Stop, or 132Stop), or an empty (Control) expression plasmids. Molecular weight markers are shown on the right in kDa. Black arrows denote expected molecular weight of truncated or full-length ORF1 proteins generated from their respective expression plasmids. **(C)** Results of the wild-type human Neo-tagged L1 retrotransposition in HeLa cells constitutively expressing functional wild-type L1 or full-length wild-type L1s containing stop codons in their ORF1 sequence. Total number of G418 resistant colonies resulting from L1 retrotransposition was divided by the number of G418 resistant colonies resulting from the transfection efficiency control (transfection with the pIRES plasmid containing neomycin resistance) was normalized to the Control/pIRES ratio for each independent experiment. Asterisk (*) denotes statistical significance between the Control and L1 (n=3, t-test, $p<0.05$).

4.4.9: Full-length wt human L1 elements containing stop codons in their ORF1 sequence suppress human L1 retrotransposition in *trans*

Based on the bioinformatic analysis of human L1 loci, we generated human L1.3 expression plasmids containing mutations that introduce stop codons at positions corresponding to amino acid 110 (110Stop), 130 (130Stop) or 132 (132Stop) of the human ORF1p (Figure 7A). A previously reported L1.3 plasmid containing a stop codon at position 119 of the ORF1p (119Stop) (*7*) was also included in the study (Figure 7A). Western blot analysis with anti-ORF1p polyclonal antibodies demonstrated that all L1 stop constructs expressed detectable levels of truncated ORF1p with expected molecular weights (Figure 7B). The 110Stop L1 produced the lowest steady-state levels of truncate ORF1 protein (Figure 7B). We next tested the effect of these L1 elements on L1 retrotransposition using transient cotransfections of HeLa cells. Consistent with the expression result, the 110 L1Stop construct had no significant impact on L1 retrotransposition (Figure 7C). Consistent with the previously published data (*7*) the 119Stop construct reduced L1 retrotransposition to about 80% of the control. The 130Stop and 132Stop containing L1 elements had the most suppressive effect on retrotransposition of the cotransfected human L1Neo in HeLa cells (Figure 7C). L1 was previously shown to have an antisense promoter (*127, 128*) and to express ORF0 protein (*128*). A plasmid supporting luciferase expression driven by the L1 5'UTR was used to control for a potential effect of the antisense mRNA generated by the 5’ UTR (*105, 129*). Cotransfection of the L1Neo plasmid with this L1 5’UTR construct demonstrated that the *trans* L1 5' UTR had no negative effect on L1 mobilization (Figure 7C).

4.4.10: wt L1 retrotransposition is suppressed in HeLa cells constitutively expressing full-length wt L1 elements containing premature stop codons in their ORF1 sequence

To assess the effect of truncated ORF1 proteins on L1 retrotransposition in a more biologically relevant manner, we generated HeLa cell lines each harboring either untagged functional human L1 (hL1wt) or one of the above-described L1 Stop elements (119Stop, 130Stop, or 132 Stop) (Figure 8A). All constitutively expressed truncated ORF1 proteins were detected at low levels in these engineered HeLa cells using western blot analysis with anti-hORF1p antibodies (Figure 8B). To assess L1 retrotransposition, human L1Neo plasmid was transiently transfected into each HeLa cell line (harboring a control plasmid, the L1 Stop or wt L1 elements) (Figure 8C). Equal transfection efficiency for each cell line was confirmed using pIRES, a neomycin expressing plasmid, that was transiently transfected into each engineered HeLa cell line under the same experimental conditions (Figure 8C, transfection efficiency control row). All engineered HeLa cell lines harboring the L1 Stop constructs supported less L1 retrotransposition relative to the control HeLa line (Figure 8C). This effect depended on the expression of truncated ORF1 proteins because the suppressive effect was reversed when the expression of truncated ORF1 proteins was lost in these engineered cell lines (Supplementary Figure S16A and S16B).

4.5: DISCUSSION

In human genomes, retrotranspositionally-incompetent full-length L1 loci outnumber functional L1 loci by about 50-fold (*10, 23, 66*). Our bioinformatic analysis of the human genome identified 1244 full-length human L1 loci containing a stop codon in their ORF1 sequence (Figure 6A and Supplementary Figure S10). Many of these L1 loci (59%) contain stop codons corresponding to amino acid positions 49, 98, 110 or 130 of the ORF1p. In contrast, the majority of mouse L1s contain a stop codon corresponding to the amino acid position 251 of the mouse ORF1p. This bias is most likely due to the presence of CpGs which results in C to T mutations caused by spontaneous deamination. Regardless of the mechanism underlying the origin of these stop codons, these results demonstrate that the human and mouse genomes contain L1 loci that, if expressed, have the potential to generate truncated ORF1 proteins.

Despite their inability to further propagation in the host genome, many inactive L1s may retain their ability to express because they contain a functional promoter present in the L1 5' UTR (Supplementary Figure S13 and Supplementary Tables 1 and 2). It has been reported that L1 mRNAs mapping to retrotranspositionally incompetent L1 loci were detected in human cells (*117, 118*). However, the approaches utilized to identify these mRNAs are limited in their ability to discriminate between authentic L1 mRNA and L1 sequences included in cellular transcripts (*130*). Analysis of recently published data sets generated by strand-specific paired-end RNAseq and 5'RACE/Pacbio approaches that unambiguously identify individual expressed human L1 loci determined that multiple L1 loci containing stop codons at different positions of their ORF1 sequence are expressed in human cells (*8*). These L1 loci are not only expressed in HEK 293 and HeLa

cells, but their mRNA they produce is likely translated because it is detected in the polyribosomal fraction of analyzed cells (*5, 8, 30*). Western blot analysis of the polyribosomal fraction extracted from HEK 293 cells demonstrated that it contained a band consistent with the size of the full-length ORF1 protein (Supplementary Figure S14). This band was not detected in HeLa cells most likely because of the difference in the levels of endogenous L1 expression detected between the two cell lines with HEK 293 cells supporting higher expression of L1Hs L1s (*8*). The same approach did not detect any specific bands consistent with the presence of truncated ORF1p proteins in either HEK 293 or HeLa cells. Manual analysis of the protein sequence encoded by the expressed L1 loci containing stop codons in their ORF1 sequence demonstrated that the majority of them contain one or more amino acid substitutions in the epitope sequence (9-22 aa of L1PA1 ORF1) recognized by our anti-ORF1p antibodies. Additionally, because expressed L1 loci harbor stop codons at different positions within their ORF1 sequences, truncated proteins generated from different L1 mRNAs would not produce a cumulative signal like the one generated by the full-length ORF1 proteins translated from different L1 transcripts. Finally, some truncated ORF1 proteins exhibit lower steady state levels than others and the full-length ORF1p (Figures 3B, 5B, 6B and 7B). Combined, these data demonstrate that coexpression of functional L1 loci and those L1 loci containing stop codons in their ORF1 sequence occurs in mammalian cells. Further, the band detected with anti-ORF1p antibodies in HEK 293 is consistent with endogenous L1 ORF1p expression, which varies between human cell lines analyzed here. This observation is consistent with the previous finding that various levels of ORF1p expression have been detected in cancerous and normal human tissues using

immunohistochemistry (*54-58, 131*) and human cell lines using western blot analysis (*132*).

The finding that non-functional L1 loci (including those containing stop codons in the ORF1 sequence) are expressed in mammalian cells along with functional L1 loci raises questions regarding the impact of this coexpression on L1 retrotransposition. It has been previously reported that retrotransposition of non-functional L1 loci is not efficiently rescued by functional L1 loci due to *cis* preference (*5, 36*). However, both L1 proteins can function in *trans* of their parental mRNA as demonstrated by retrotransposition of SINEs and SVA elements that utilize L1 proteins (*14, 15, 62*). Although, ORF1p molecules produced from different mRNAs can assemble heterotrimers (*9*), it is not known whether non-functional L1 loci, specifically those that may produce truncated ORF1 proteins, have an impact on retrotransposition of active L1 elements. The conventional thinking assumes that L1 mRNA and/or ORF1 protein levels serve as a direct indicator of L1 retrotransposition (*54, 55, 114, 115*). Consistent with this line of thinking, several examples of an increase in L1 mRNA and protein expression and corresponding upregulation in L1 retrotransposition upon elimination of suppressors of L1 expression have been reported (*48, 49, 77, 90-93, 95, 104, 133-137*). Further supporting a correlation between ORF1p expression and L1 retrotransposition, codon optimization of the mouse L1 dramatically (100-fold) increased its mRNA, protein expression, and retrotransposition (*84*). Deviating from this straight forward relationship is the observation that codon optimization of the human L1s significantly increased its mRNA and protein expression, but this increase resulted in a modest 3-5 fold increase in the human L1 mobilization (*1, 78, 84, 116*). Additionally, recently published data suggest

that the ratio of ORF1/ORF2 proteins may influence L1 retrotransposition (*49*). Therefore, the expression levels of ORF1p may not always directly correlate with L1 retrotransposition. A relationship between the levels of endogenous L1 (mRNA or ORF1p) expression and L1 retrotransposition *in vivo* is further complicated by the fact that typically the total levels of these molecules are produced by multiple expressed L1 loci. Some of these L1 loci are functional and some are defective. Therefore, a specific spectrum of expressed functional and non-functional L1 loci contributing to the total mRNA and protein signal may be important for understanding the resulting impact on the host genome from active L1 loci. Our data suggest the possibility that the levels of the ORF1p detected in any given tissue may not always directly correspond to the amount of L1 retrotransposition, and that an inverse relationship between L1 retrotransposition and the relative amount of the total ORF1p may exist. This scenario is based on the observations that both the full-length and C-terminally truncated ORF1 proteins suppress retrotransposition of active human L1 in transient assays (Figures 3C, 5C, 6C, 7C and Supplementary Figure S7), as well as in engineered cells harboring the full-length wt L1 or L1s containing stop codons in their ORF1 sequence (Figure 8C). These data support that expression of non-functional L1 loci containing stop codons in their ORF1 sequence may be beneficial to the host genome as it may reduce accumulation of *de novo* L1 inserts.

While both the full-length and the truncated ORF1 proteins suppress human L1 retrotransposition when supplied in *trans,* the underlying mechanisms of their effects are likely different. One piece of experimental evidence that supports this likelihood is that the full-length human and mouse ORF1 proteins have contrasting effects on

retrotransposition of their respective elements. While the full-length human ORF1p suppressed mobilization of the human L1 and chimeric L1s containing human ORF1 sequence, mouse full-length ORF1p increased retrotransposition of the codon-optimized mouse L1 and the chimeric L1s containing mouse ORF2p in NIH 3T3 cells (Figure 2). This differential effect of the mouse and human ORF1 proteins on L1 mobilization could be a factor contributing to the differences in retrotransposition rates of mouse and human L1 elements containing codon-optimized wild-type sequences (*1, 78, 116*). The suppressive *trans* effect of the human ORF1p on human L1 retrotransposition is also opposite from its stimulatory impact of Alu mobilization (*15*). Although understanding the underlying mechanism and species-specific differences of the full-length ORF1 effect on L1 retrotransposition is beyond the scope of this manuscript, our results are consistent with the recent observation that ORF1p/ORF2p ratio may influence human L1 integration (*49*). In conjunction with previously published results, our data regarding the *trans* effect of the full-length ORF1p on L1 retrotransposition are consistent with the proposed hypothesis that the ORF1p content within L1 RNPs may impact retrotransposition (*49*).

In contrast to the opposing effects of the full-length ORF1 proteins on human and mouse L1 retrotransposition, truncated ORF1 proteins from both species suppress mobilization of their respective full-length L1 elements (Figures 3C, 5C, 6C, 7C, Supplementary Figure S7B and S11D). Our results demonstrate that the minimal unit of the human ORF1p that is able to suppress L1 retrotransposition in *trans* consists of the N-terminal and C-C domains. The effect of truncated ORF1 proteins is consistent with a dominant negative mechanism of their action, as it depends on the ability of these proteins to form heterodimers with the full-length ORF1p (Figure 3D and

Supplementary Figure S15). Genetically engineered disruption of the coiled-coil domain function abolished the ability of truncated ORF1p to form heterodimeric complexes with the full-length ORF1p. These changes also eliminated the suppressive effect of the 132 protein on L1 retrotransposition (Figure 4B and 4E). The fact that longer truncated human ORF1 proteins are less efficient at suppressing L1 retrotransposition than those containing only the N-terminus and C-C domain is also consistent with the dominant negative mechanism of their interference. Although it is not known whether the truncated ORF1 proteins could be incorporated into L1 RNPs and interfere with L1 integration, our data demonstrate that the full-length ORF1p can ferry truncated wild-type 132 protein into the nucleus (Figure 4D, hORF1+132). In contrast, the coiled-coil domain mutant 132 protein remained cytoplasmic in the presence of the full-length ORF1p (Figure 4D, hORF1+132M). This result also demonstrates that the full-length and truncated ORF1 proteins have different subcellular localization. This may explain some observed variability in localization of the ORF1p signal detected by immunohistochemistry in human samples (*54, 55, 58, 103*), if the ORF1p-specific antibodies can recognize both the full-length and truncated ORF1 proteins generated by expressed L1 loci. This result also supports that the RRM/C-terminal portion of the protein is important for its nuclear localization and that having two (or potentially one) full-length ORF1 molecules within the trimer may be sufficient for gaining access to the nucleus.

The experimental evidence presented here supports that the efficiency of the dominant negative effect of the truncated L1 ORF1p proteins on retrotransposition is

influenced by the levels of their expression and the extent of their truncation (Figures 3B, 3D, 5B, 5C, 6B, 6C, 7B, 7C and Supplementary Figure S7B). These findings have important biological implications, if the rules we identified here apply *in vivo*. Analysis of endogenously expressed L1 loci harboring stop codons in their ORF1 sequence showed that their spectrum varies among different human cell lines (HeLa versus HEK 293) (Supplementary Tables 1 and 2). The expression levels, positions of stop codons, and the composition of specific expressed L1 loci are therefore expected to influence the resulting amount of DNA damage generated from endogenous active L1s. For example, based on the results obtained using transient transfections of L1 loci with stop codons roughly corresponding to amino acid positions 110-130 of the human ORF1p may be optimal for guarding against accumulation of *de novo* L1 events. However, generation of engineered HeLa cells stably expressing different L1Stop elements supports that, even at low levels, truncated ORF1 proteins are able to suppress L1 retrotransposition when constitutively expressed in human cells (Figure 8). This outcome could be pointing to some biologically relevant differences related to the ORF1p life cycle during transient versus constitutive expression. It is worth noting that our findings demonstrate that the negative effect of L1 elements containing stop codons in their ORF1 sequence on L1 retrotransposition is reversible because the loss of constitutive expression of the full-length or truncated ORF1 proteins lead to the loss of suppression of L1 retrotransposition (Supplementary Figure S16A and S16B). This finding suggests that the protective effect from expression of L1s with ORF1 stop codons could be gained or lost following epigenetic changes due to aging, environmental exposures, or cellular differentiation or transformation. Additionally, these results combined with the recently

reported cell-type specific expression of L1 loci (*8*) suggest that the protective effect of L1 stop loci may be tissue specific. Furthermore, the impact of L1Stop loci on L1 retotransposition may be influenced by the fact that the presence of some of these stop codons within fixed L1 loci is polymorphic in the human population (Supplementary Figure S9).

Numerous cellular factors and pathways have been identified to suppress L1 expression and integration, establishing the existence of significant redundancy in mechanisms downregulating L1 retrotransposition (*48, 49, 77, 90-93, 95, 104, 133-137*). The suppressive *trans* effect of hORF1p on human L1 retrotransposition is another potential mechanism that may influence L1-induced damage *in vivo*. Our findings support that coexpression of retrotranspositionally incompetent L1s containing stop codons in their ORF1 sequence with functional L1 loci may slow down accumulation of *de novo* L1 integration events. Our data, along with other recent reports (*49, 116, 138*), suggest that the relationship between detected ORF1p levels may not always be proportional to the extent of L1 retrotransposition, and that an inverse correlation between ORF1p expression levels and L1 retrotransposition may exist under certain circumstances. The factors influencing L1 retrotransposition in any given cell may include the relative ratios of the functional full-length ORF1p and truncated proteins, as well as the length of truncated ORF1 proteins expressed in a given cell.

CHAPTER 5: DISCUSSION

In 2011, very little was known about L1 proteins in the context of the mammalian cellular environment. The majority of the knowledge about L1 proteins was acquired through *in vitro* methods using bacteria and/or baculovirus purified L1 proteins or their fragments (*3, 4, 12, 16, 17, 31, 33, 35, 37-43*). We adapted non-reducing western blot analysis and the mammalian two-hybrid system to assay ORF1p self-interaction in cultured mammalian cells (Chapter 3). Using these approaches, we have discovered that ORF1 proteins produced from different mRNAs readily interact in *trans* (Chapter 3). These ORF1p *trans* interactions could have important biological impact on the L1 replication cycle. Some mammalian genomes harbor hundreds of full-length L1s that contain only ORF1 sequence (half L1s) suggesting a possible benefit for the retention of such loci (*139-141*). In addition, the splicing and/or premature polyadenylation of L1 mRNAs could potentially produce functional and/or non-functional ORF1 proteins in normal tissues (*77, 129, 142*). These ORF1 proteins could have an impact on L1 retrotransposition in *trans*.

A human genome harbors many full-length L1 loci with retrotranspositionally-incompetent L1 loci greatly outnumbering functional L1s (about 5000 versus 100) (*10, 23, 66*). Many of the retrotranspositionally-incompetent L1s retain their expression potential (*5, 36*). In fact, it has been reported that RNA from a subset of these loci is detected in human cells (*60, 118, 143*). Furthermore, experimental evidence supports that the expression profiles of L1 loci may vary from tissue to tissue suggesting that different cells or tissues may have unique expression profiles of functional and non-

functional L1 loci (*144*). Our bioinformatic analysis of full-length human and mouse L1 loci identified 1244 human and 198 mouse non-functional L1 loci containing at least one stop codon in their ORF1 sequence (Chapter 4, Figure 6A and Figure S4.10). Interestingly, certain positions within ORF1 were highly over-represented even though about 40% of the ORF1 sequence is only one base pair mutation away from becoming a stop codon (Figure S4.8). These results suggest that the human and mouse genomes contain a subset of L1 loci that, if expressed, have the potential to generate truncated ORF1 proteins. Previously published results demonstrated that some of these L1 loci are indeed expressed in human cells (*60, 118*). While it has been reported that functional L1 loci do not efficiently rescue retrotransposition of non-functional L1 loci due to *cis* preference (*5, 36*), it is not known whether non-functional L1 loci have an impact on retrotransposition of functional L1 loci. It is well established that Alu and SVA elements can utilize both L1 proteins in *trans* to retrotranspose in to a new genomic location (*14, 15, 62*). Additionally, even though Alu elements do not require L1 ORF1p for their mobilization, transient expression of this protein *in trans* improves Alu retrotransposition (*15*). Of the two proteins encoded by L1, ORF1p is far more abundant than the ORF2p due to the bicistronic nature of L1 mRNA and the non-conventional translation of the ORF2 protein (*21*). Various levels of the ORF1p expression have been detected in human tumor and normal tissues using immunohistochemistry (*54, 55, 58*). However, the relevance of endogenous ORF1p expression to L1 retrotransposition remains unknown, because we do not fully understand whether coexpression of non-functional L1 loci has any impact on the mobilization of active L1 elements.

To gain some understanding of the interplay between functional and non-functional L1 loci, we tested whether ORF1p expression affects L1 retrotransposition *in trans*. In contrast to its positive effect on Alu mobilization, expression of the full-length hORF1p suppressed human L1 retrotransposition in human and mouse cells (Chapter 4, Figure 1). Mouse ORF1p did not have any negative effect on human L1 mobilization in either human or mouse cells (Chapter 4, Figure 1). Utilization of chimeric L1 constructs suggested that the observed negative effect involves the L1 ORF1p, as transient expression of human ORF1p only suppressed retrotransposition of L1 elements containing human ORF1 sequence regardless whether the experiments were carried out in human or mouse cells (Chapter 4, Figure 2). This result is consistent with our previous observation that human ORF1p molecules expressed from different mRNAs can associate in *trans*, while no *trans* association was detected between the human and mouse ORF1 proteins (*9*). By generating expression plasmids designed to express truncated ORF1 proteins, we determined that both the N-terminus and the coiled-coil domains of the human ORF1p are required for this suppressive effect (Chapter 4, Figure 3C). This conclusion was derived from the fact that the coiled-coil domain of ORF1p did not suppress L1 mobilization by itself and that the N-terminal portion of the protein was unstable. Additionally, despite being expressed, human mouse chimeric N C-C proteins did not suppress L1 mobilization (Figure S4.4). Consistent with our previous observation (*9*), truncated ORF1 proteins that suppressed L1 mobilization were able to form heterocomplexes with the full-length ORF1 protein. Point mutations in the ORF1p leucine-zipper motif designed to disrupt the function of the coiled-coil domain eliminated detection of heterodimeric complexes by western blot analysis and abolished

the suppressive effect on L1 retrotransposition (Chapter 4, Figure 4B and 4E). Furthermore, using nuclear/cytoplasmic fractionation of transfected cells, we determined that the truncated 132 protein was ferried by the full-length ORF1p from the cytoplasm into the nucleus (Chapter 4, Figure 4D). In contrast, the leucine zipper mutant 132 protein remained cytoplasmic in the presence of the full-length ORF1p. These findings suggest that heterotrimers containing truncated ORF1 proteins could be incorporated into L1 RNPs and interfere with L1 integration by compromising ORF1p trimer function. Similar to our findings with the truncated human ORF1 proteins, a truncated mouse ORF1 protein also suppressed retrotransposition of the mouse (but not human) L1 (Figures S4.11 and S4.12).

Mammalian genomes are known to harbor an excess of retrotranspositionally incompetent L1 loci in comparison to functional loci. Specifically, we identified 1244 full-length human L1 loci containing a stop codon in the ORF1 sequence. The majority of these human L1 loci (59%) contained premature stop codons at specific positions (48, 98, 110 and 130) (Chapter 4, Figure 6A). The most common position within the 198 mouse L1 loci that were analyzed was at amino acid position 251 (Figure S4.10). Our data support that, if transcribed, human L1 loci containing stop codons at positions 48 or 98 of the ORF1 protein would be expected to produce unstable truncated ORF1 proteins. We confirmed that, if transcribed, the other dominant L1 loci could generate stable truncated ORF1 proteins. One of these constructs 119 was previously reported (*7*). The 119 construct produced detectable levels of truncated ORF1 protein (Chapter 4, Figure 7B) and, as previously reported, reduced L1 retrotransposition by about 20% when transiently cotransfected with the Neo-tagged L1 (Chapter 4, Figure 7C). Full-length

L1s producing ORF1 proteins truncated at amino acids 130 and 132 suppressed L1 mobilization by about 50-60% (Chapter 4, Figure 7C). Chronic expression of stop codon-containing L1s had a more significant suppressive effect on L1 retrotransposition. This 80% suppression was similar to the suppressive effect of the chronically expressed fully-functional human L1 (Chapter 4, Figure 8C). The suppressive effect was reversible as the loss of constitutive expression of the full-length or truncated ORF1 proteins coincided with the loss of suppression of L1 retrotransposition (Figure S4.16). These findings support that coexpression of retrotranspositionally incompetent and functional L1 loci may protect the host genome from the accumulation of *de novo* L1 integration events.

Interestingly, the full-length human and mouse ORF1 proteins had contrasting effects on retrotransposition of their respective L1s. While the full-length human ORF1p suppressed mobilization of L1s containing human ORF1 sequence, mouse full-length ORF1p increased retrotransposition of L1s containing mouse ORF2p (Chapter 4, Figure 2). This effect was specific to mouse cells (NIH 3T3). Even though understanding the underlying mechanism of this effect is beyond the scope of this manuscript, it provokes speculations relevant to the biology of mouse and human elements. Our observations are consistent with results obtained with codon-optimization of the human and mouse L1 elements. As expected, codon optimization significantly increased L1 mRNA and protein levels of both elements. This change translated into a greater than 100-fold increase in the retrotransposition of the mouse element (*84*), but only a modest 3-5-fold boost in the human L1 mobilization (*1, 78, 116*). This finding suggested that even though mouse and human elements share many steps in their integration process they

may have different requirements for the abundance of their proteins. Our results are also consistent with the recent observation that reduced levels of ORF1p may favor human L1 integration (*49*). The opposing strategies of utilizing ORF1p by the human and mouse L1 elements may potentially be rooted in the evolutionary differences in their cellular environment. Human L1 evolution has involved the rise and fall of a single active L1 lineage (*44, 145, 146*). In contrast, mouse genomes contain several families of concurrently active L1 elements (*147-150*). Our data suggest a possibility that the presence of multiple active L1 subfamilies dictated the development of the positive influence of the mouse ORF1p on mouse L1 retrotransposition. Alternatively, this property of the mouse ORF1 protein could have permitted the coexistence of more than one active L1 subfamily in the mouse genome.

We developed monoclonal antibodies against the human endonuclease domain of the ORF2p, which allows our field to detect untagged ORF2p generated from a full-length L1 construct (Chapter 2) (US Patent Pending 14/943,942). Our findings have established that ORF1p generated from different mRNA molecules efficiently interact in *trans* (Chapter 3). These findings contested the existing dogma that L1 ORF1 protein is not involved in *trans* interactions during L1 retrotransposition. The antibody partially inhibits ORF2 endonuclease activity *in vitro*, and the antibody has promising future applications *in vivo*. Based on these initial findings, we identified a novel mechanism of ORF1p-associated suppression of L1 retrotransposition mediated through the *trans* interactions of ORF1 proteins produced by functional and defective L1 loci (Chapter 4). These findings have a substantial contribution to our understanding of L1 biology relevant to L1 amplification in its host. Our findings also have important biological

relevance because recent studies have demonstrated that both functional and non-functional L1 loci are coexpressed in different cell types (*60, 61, 143*) suggesting that functional and non-functional ORF1p could be coexpressed in the same cell. These results combined with our findings support that truncated and full-length ORF1p may be negatively impacting L1 retrotransposition in *vivo* as they do in cultured cells. We have also identified the presence of polymorphic stop codons within the ORF1 sequence of fixed human L1 loci (Figure S4.9). This finding suggests an intriguing possibility that human genomes may experience different levels of L1 retrotransposition based on the presence of L1 loci containing polymorphic stop codons within the ORF1 sequence or their potential tissue-specific expression.

APPENDIX 1: Information Supplemental to "Chapter 2: Development of a monoclonal antibody specific to the endonuclease domain of the human LINE-1 ORF2 protein"

Figure S2.1

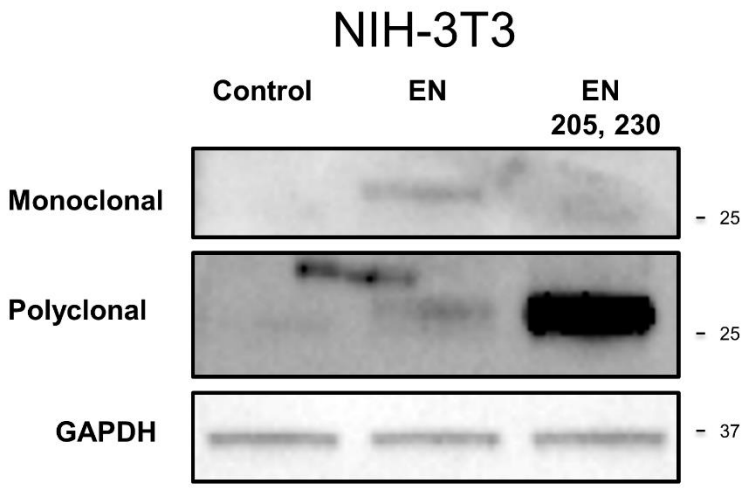

Figure S2.1. Analysis of expression of functional and non-functional human ORF2 endonuclease domains in NIH-3T3 cells. Western blot analysis of proteins generated from expression plasmids containing codon-optimized functional (EN) and non-functional (EN 205,230) ORF2 endonuclease sequences transiently transfected in NIH-3T3 cells with anti-human ORF2p monoclonal antibody **(top)**, anti-human ORF2p polyclonal antibodies **(middle)**, or GAPDH **(bottom).** Control lane indicates cells transiently transfected with an empty vector. 25 and 37 kDa are molecular markers.

Figure S2.2

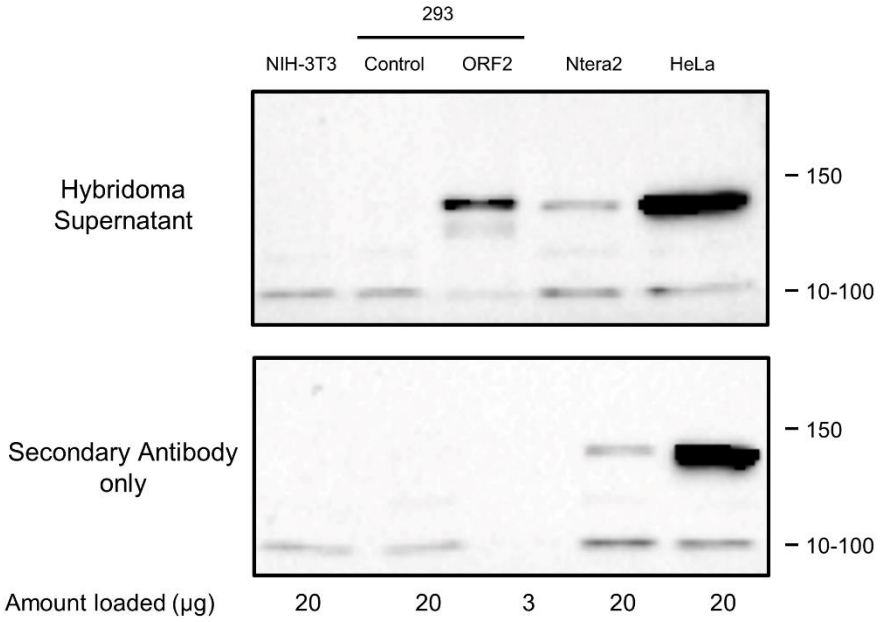

Figure S2.2. Analysis of endogenous ORF2p in different cell lines. (top) Western blot analysis of total cell lysate from the following cell lines: NIH-3T3, 293, Ntera2 and HeLa using hybridoma supernatant. Protein lysate from 293 cells transiently transfected with an expression plasmid containing codon-optimized ORF2 was used as a positive control for ORF2p expression (third lane). Control lane indicates 293 cells transiently transfected with an empty vector. Positions of molecular markers are indicated on the right as 100 or 150 kDa. **(bottom)** The same experiment and analysis as in (top), but using secondary antibodies only. Positions of molecular markers are indicated on the right as 100 or 150 kDa. Total amount of cell lysate loaded is in micrograms (µg).

Figure S2.3

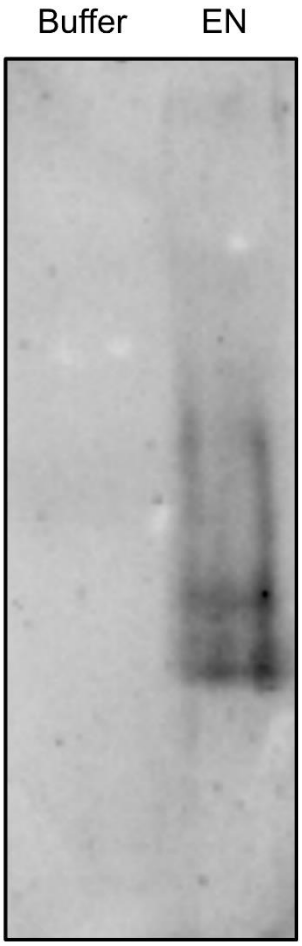

Figure S2.3 Analysis of functional ORF2p endonuclease in native conformation. Western blot analysis of the bacterially purified ORF2p endonuclease (EN) fractionated under native conditions using hybridoma supernatant. Buffer lane indicates storage buffer used for purified ORF2p endonuclease.

Figure S2.4

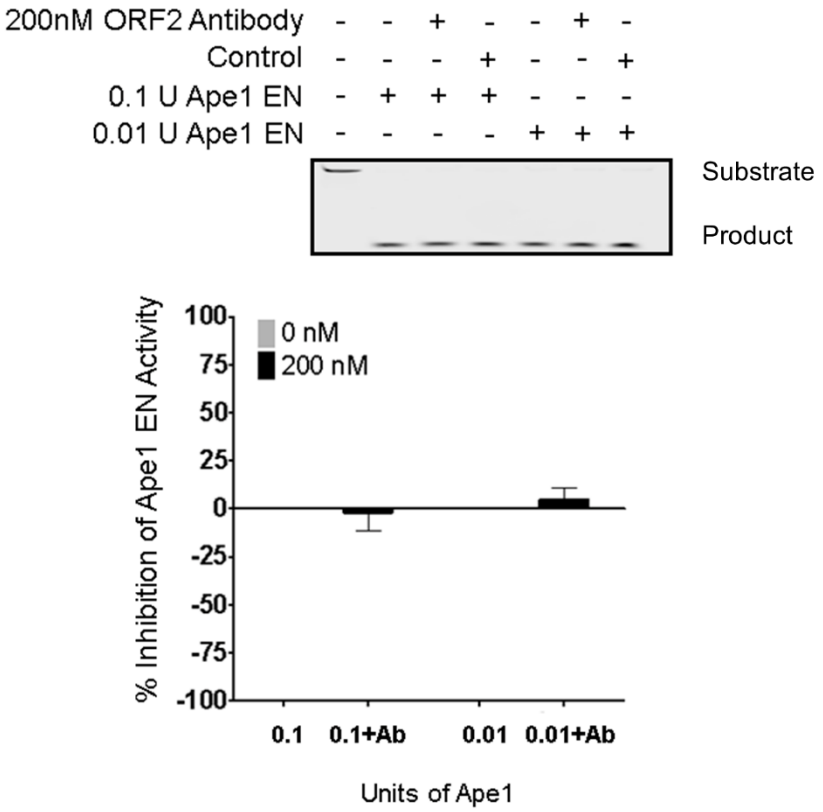

Figure S2.4. Monoclonal anti-ORF2p antibody does not inhibit the APE1 endonuclease activity *in vitro*. (top) SDS-PAGE analysis of *in vitro* APE1 endonuclease cleavage assay using monoclonal anti-ORF2 antibody. ORF2 antibody denotes the addition of 200nM of the monoclonal anti-ORF2p antibody, control indicates the addition of the same volume of the buffer used for the reactions containing monoclonal anti-ORF2 antibody, APE1 denotes bacterially purified human APE1 endonuclease, 0.1 and 0.01 units of APE1 were tested. **(bottom)** Quantitation of the results of the *in vitro* APE1 endonuclease cleavage assay in A. Results were normalized to 0nM control (N=3). Equation used to determine percent (%) inhibition is listed in the materials and methods section.

Figure S2.5

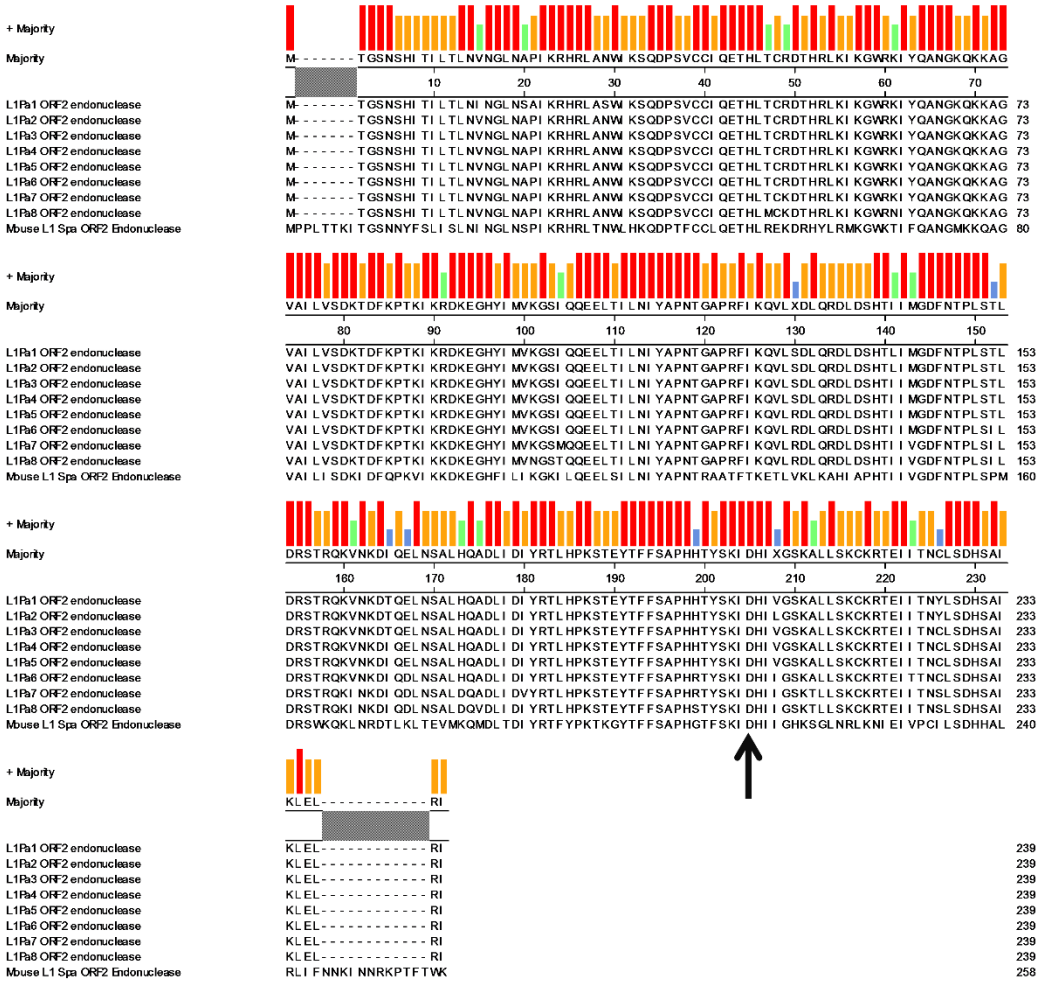

Figure S2.5. Analysis of ORF2p endonuclease conservation in human and mouse. Alignment of ORF2p endonucleases of L1Pa families in humans and the ORF2p endonuclease domain of mouse L1 Spa. Black arrow indicates area of the epitope of anti-ORF2p monoclonal antibody.

APPENDICES

APPENDIX 2: Information Supplemental to "Chapter 3: Characterization of L1 ORF1p self-interaction and cellular localization using a mammalian two-hybrid system"

Figure S3.1

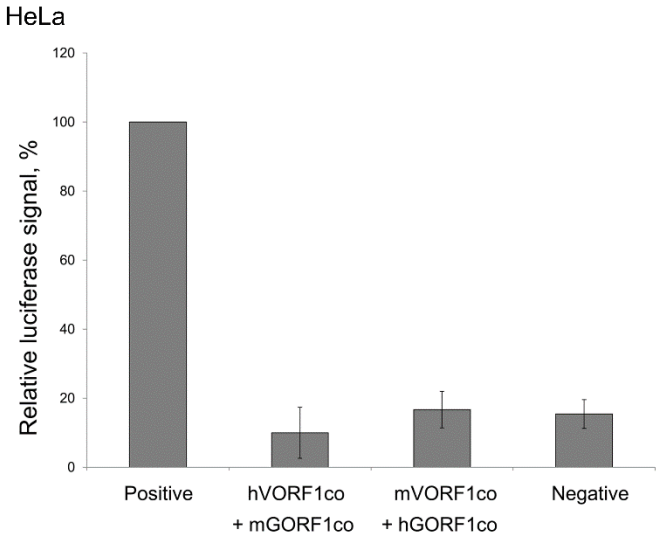

Figure S3.1. Analysis of human and mouse ORF1p self-interaction using the M2H system. M2H results for hVORF1co + mGORF1co and mVORF1co + hGORF1co interaction in HeLa cells normalized to positive control. No interaction is detected between these protein pairs (T-test, p-value >0.05).

Figure S3.2

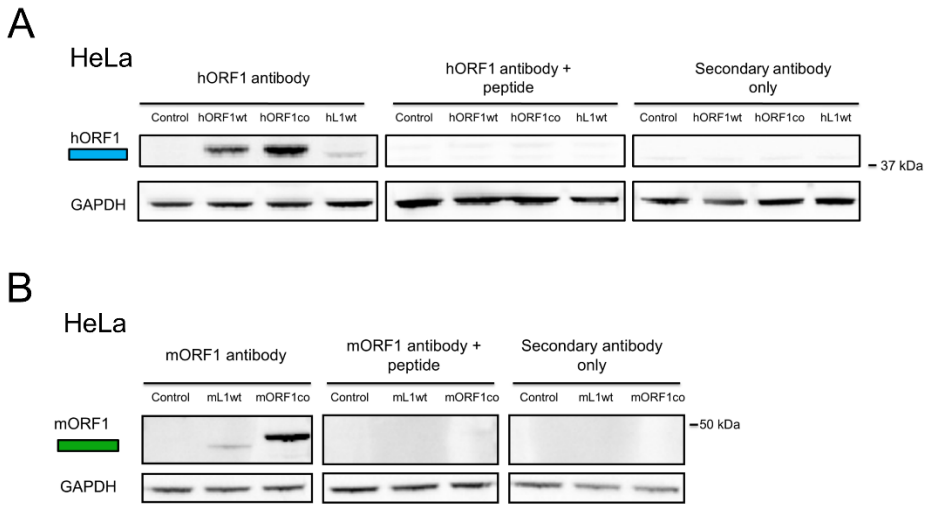

Figure S3.2. Validation of human- and mouse-specific ORF1p antibodies. A. Human ORF1 polyclonal antibodies were generated against a peptide within the N-terminus of the hORF1p. Western blot analysis of total cellular lysates harvested from HeLa cells transfected with a plasmid expressing hORF1co protein (blue rectangle). (**Left to right**) First Panel: detection of hORF1p in HeLa cells (transiently transfected with an empty control vector, hORF1wt, hL1wt or hORF1co plasmids) with human ORF1p-specific antibodies; Second Panel: detection of hORF1p in HeLa cells with human ORF1p-specific antibodies and the addition of the peptide used for the antibodies' generation (for the peptide sequence, see material and methods); Third Panel: detection of hORF1p in HeLa cells with secondary antibodies alone. **B.** Mouse ORF1 polyclonal antibodies were generated against a peptide located within the C-terminus of mORF1p (green rectangle). Western blot analysis of protein lysates from HeLa cells transiently transfected with mL1 wt and mORF1co expression plasmids. mORF1co expression plasmid adds a tag to the ORF1p, causing the difference in observed molecular weight between mL1wt and mORF1co. (**Left to right**) Detection of mORF1p expressed in cells transiently transfected with mL1wt and mORF1co expression plasmids. Loss of protein detection in the presence of mORF1 peptide (for peptide sequence see material and methods) pre-incubated with mORF1 antibodies prior to use for western blot analysis. The third group is incubated with secondary goat antibodies only.

Figure S3.3

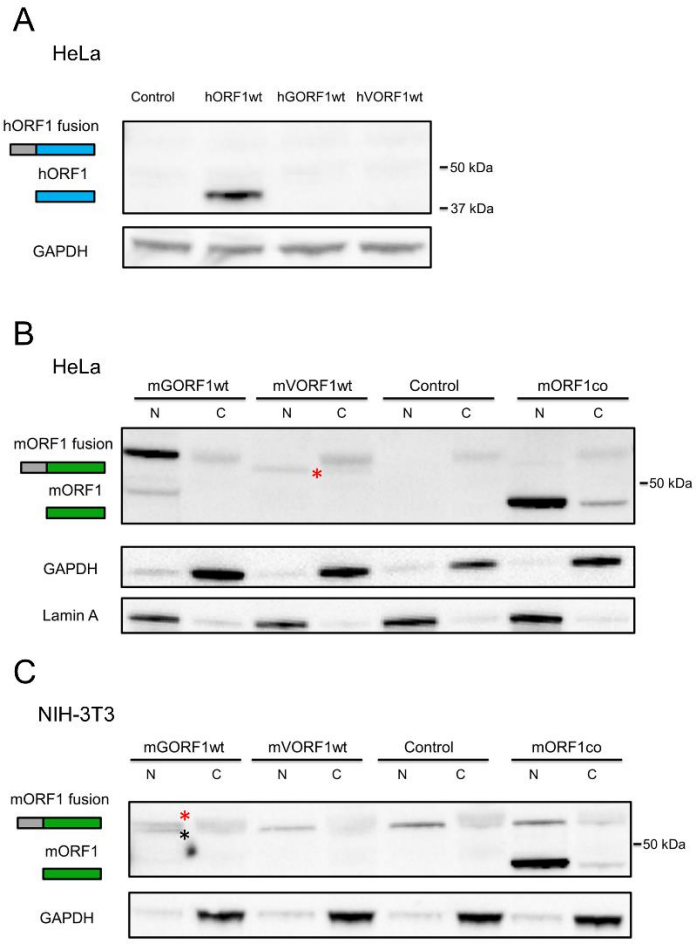

Figure S3.3. Analysis of subcellular localization of ORF1 fusion proteins expressed by plasmids containing wild type L1 sequence. A. Western blot analysis of total hORFwt (blue rectangle), hGORF1wt or hVORF1wt (grey/blue rectangle) protein transiently expressed in HeLa cells. GAPDH is used as a loading control. Control lanes indicate cells transfected with an empty vector. Only untagged hORF1wt is detected. **B.** Western blot analysis of nuclear and cytoplasmic fractions of mORF1co (green rectangle), mGORF1wt and mVORF1wt (grey/green rectangle) proteins transiently expressed in HeLa cells. Proteins in nuclear (N) and cytoplasmic (C) fractions were detected with mouse specific ORF1p polyclonal antibodies. GAPDH (cytoplasmic marker) and Lamin A (nuclear marker) are used as loading and cell fractionation controls. 50 and 75 kDa indicate molecular weights. Control lanes indicate cells transfected with an empty vector. Red asterisk indicates the mORF1p-specific band detected in the nuclear and cytoplasmic fractions. **C.** The same experiment and analysis as in B is performed in NIH-3T3 cells. Red asterisk represents mORF1p specific band, black asterisk represents a non-specific band.

Figure S3.4

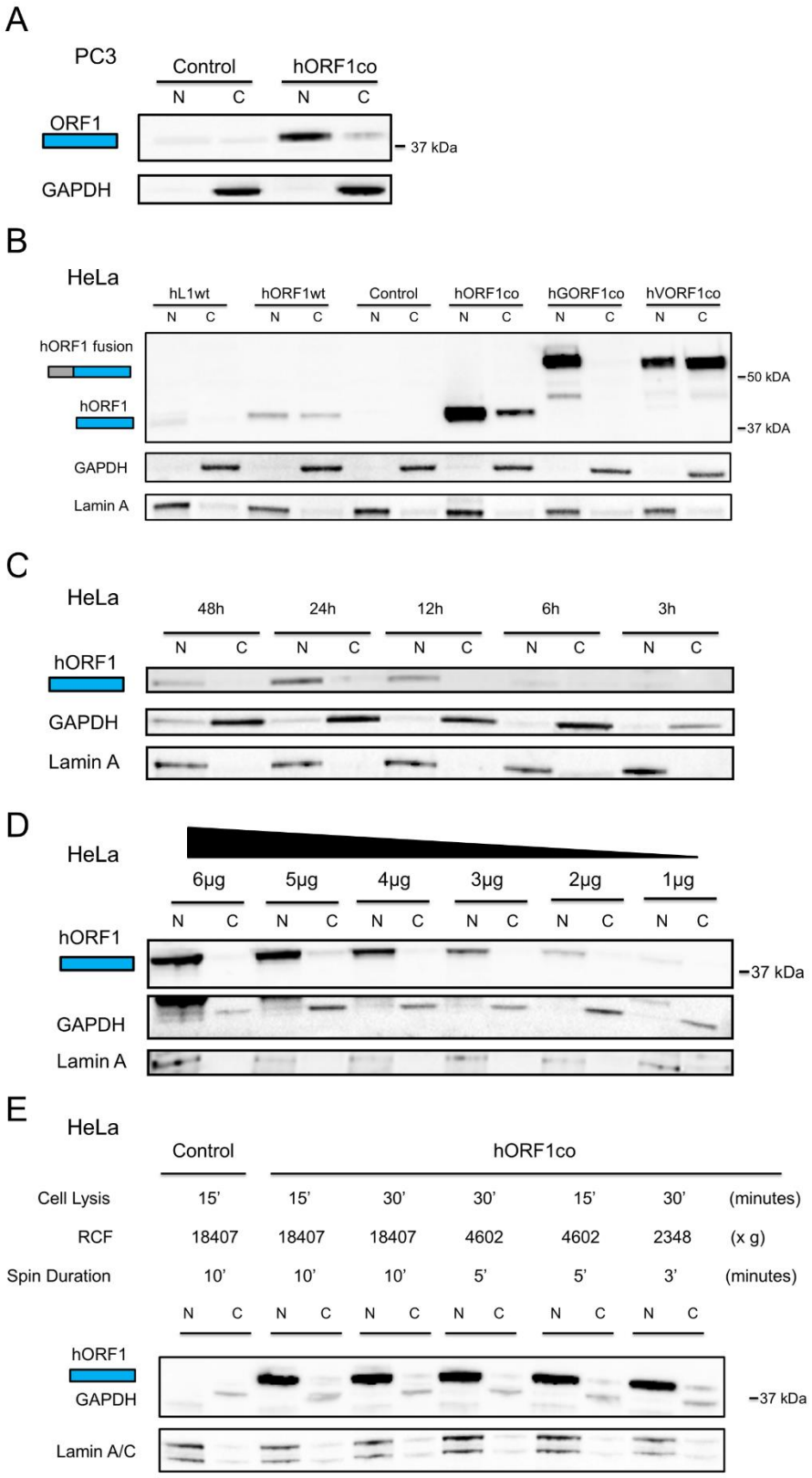
A
PC3
Control
hORF1co
N C N C
ORF1
37 kDa
GAPDH
B
HeLa
hL1wt
hORF1wt
Control
hORF1co
hGORF1co
hVORF1co
N C N C N C N C N C N C
hORF1 fusion
hORF1
50 kDA
37 kDA
GAPDH
Lamin A
C
HeLa
48h
24h
12h
6h
3h
N C N C N C N C N C
hORF1
GAPDH
Lamin A
D
HeLa
6µg
5µg
4µg
3µg
2µg
1µg
N C N C N C N C N C N C
hORF1
37 kDa
GAPDH
Lamin A
E
HeLa
Control
hORF1co
Cell Lysis 15' 15' 30' 30' 15' 30' (minutes)
RCF 18407 18407 18407 4602 4602 2348 (x g)
Spin Duration 10' 10' 10' 5' 5' 3' (minutes)
N C N C N C N C N C N C
hORF1
GAPDH
37 kDa
Lamin A/C

Figure S3.4. Effect of cell type and transfection conditions on the subcellular localization of hORF1p. A. Western blot analysis of hORF1co transiently expressed in a prostate cancer cell line (PC3). Protein is detected with human specific ORF1p polyclonal antibodies. GAPDH (cytoplasmic marker) is used as a loading and cell fractionation control. Control lanes indicate cells transfected with empty vector. **B.** Western blot analysis of hORF1co (blue rectangle), hGORF1co and hVORF1co (grey/blue rectangle), and hL1wt and hORF1wt transiently expressed in HeLa cells. Equal fractions of the total protein in nuclear and cytoplasmic extracts for each construct transiently transfected in HeLa cells were analyzed. This approach is used to take into account the difference in the total amount of protein collected from each cellular fraction. Compare these results with the results shown in Figure 4, where the analysis was done by loading the same amount of protein for each subcellular fraction. Proteins in nuclear (N) and cytoplasmic (C) fractions were detected with human ORF1p-specific polyclonal antibodies. GAPDH (cytoplasmic marker) and Lamin A (nuclear marker) are used as loading and cell fractionation controls. 50 and 75 kDa are molecular weights. Control lanes indicate cells transfected with empty vector. **C.** Time course analysis of hL1wt protein expressed in transiently transfected HeLa cells. Cells were harvested over a 48 hour period at indicated time points (h) post transfection. hORF1p in nuclear (N) and cytoplasmic (C) fractions was detected with human specific ORF1p polyclonal antibodies. GAPDH (cytoplasmic marker) and Lamin A (nuclear marker) are used as loading and cell fractionation controls. **D.** Western blot analysis of a dose curve: HeLa cells were transfected with increasing amounts of hORF1co plasmid (1 to 6 micrograms (µg)). The same loading controls and antibodies were used as in B and C. 37 kDa is a molecular weight. **E.** Multiple conditions for processing were used to rule out the effect of the duration of cell lysis and sample processing on ORF1p subcellular localization. The duration of cell lysis on ice, centrifugation time, or centrifugation speed did not have any effect on the ORF1p subcellular localization. ORF1p in nuclear (N) and cytoplasmic (C) fractions was detected with human-specific ORF1p polyclonal antibodies. GAPDH (cytoplasmic marker) and Lamin A (nuclear marker) are used as loading and cell fractionation controls. Control lanes indicate cells transfected with empty vector.

Figure S3.5

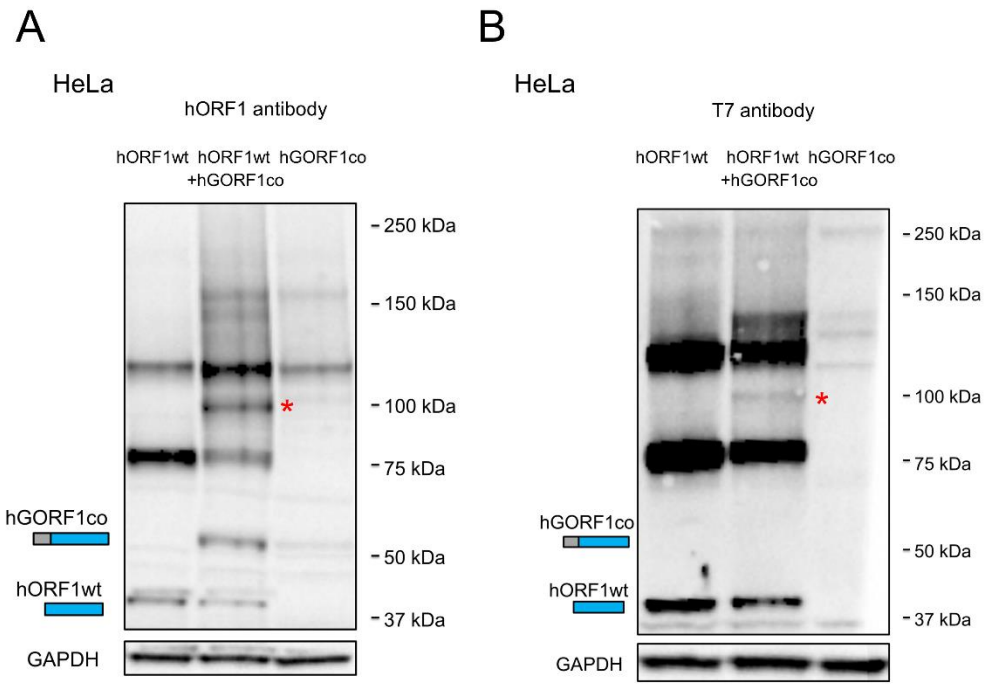

Figure S3.5. ORF1p heterodimer formation analysis. A. Western blot analysis of ORF1p expressed by hORF1wt (blue rectangle labeled hORF1wt) and hGORF1co (grey/blue rectangle labeled hGORF1co) vectors transiently transfected alone or cotransfected in HeLa cells. Red asterisk denotes an ORF1p heterodimer. Proteins are detected with human specific polyclonal antibodies. GAPDH was used as loading control. Molecular weights are indicated on the right in kilodaltons (kDa). **B.** The same experiment as in A. T7 tag-specific antibodies were used for protein detection. Red asterisk denotes an ORF1p heterodimer. GAPDH was used as loading control. Molecular weights are indicated on the right in kilodaltons (kDa).

Figure S3.6

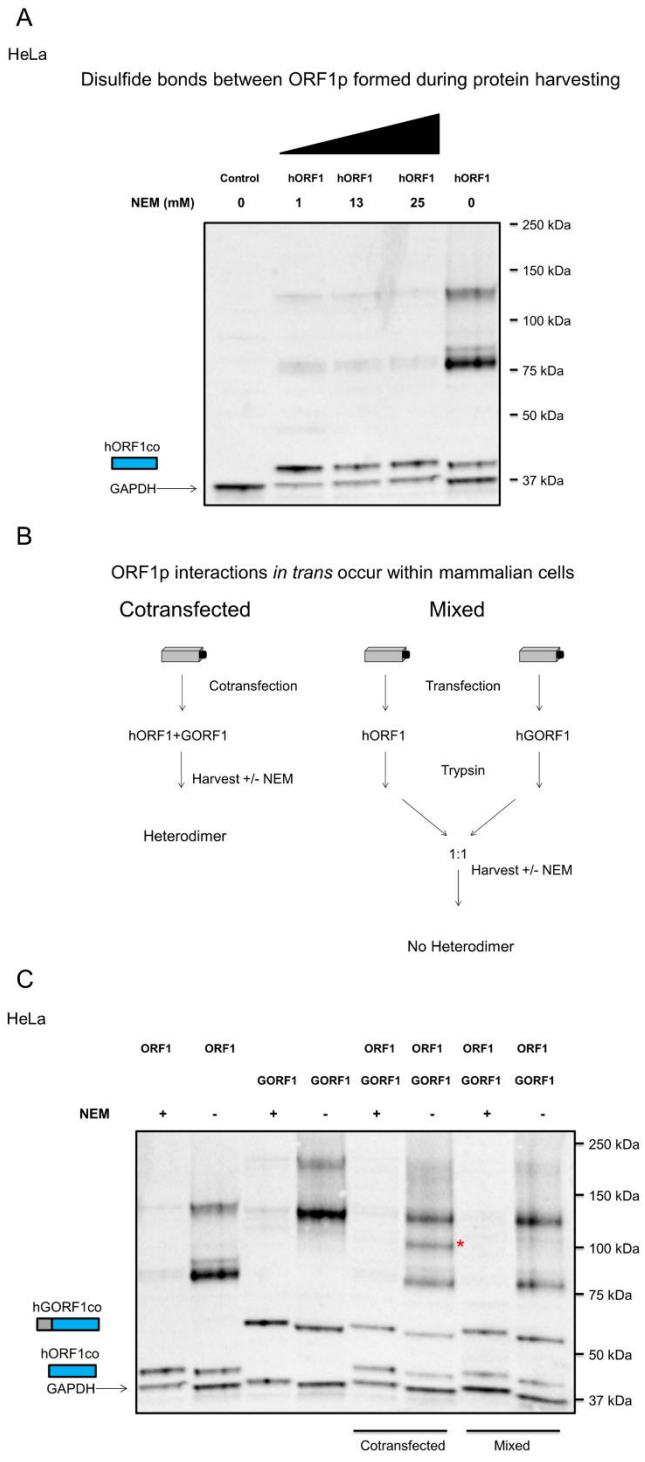

Figure S3.6. Disulfide bond formation and ORF1p self-interaction. A. Western blot analysis of hORF1co expressed in HeLa cells transiently transfected with hORF1co expression plasmid. Total protein was harvested in the presence of increasing amounts of N-ethylmaleimide (NEM) and analyzed on a non-reducing SDS-PAGE gel. Molecular weights in kilodaltons (kDa) are indicated on the right. **B.** Schematic of the experimental approach. Total protein was collected with or without NEM (25mM) from HeLa cells cotransfected with hORF1co and hGORF1co expression plasmids (see left, labeled cotransfected) or transfected individually (see right, labeled mixed). The control is designed to determine whether hORF1p and hGORF1p can interact during sample processing. Briefly, cells transfected with either hORF1 or hGORF1 expression plasmids were trypsinized and mixed in equal amounts. Total protein was extracted from the mixture with or without the addition of NEM. No interactions *in trans* were detected under these experimental conditions. **C.** Western blot of ORF1p (blue rectangle labeled hORF1co is unfused, grey/blue rectangle labeled hGORF1co is a fusion ORF1 protein) interactions *in trans* occurring within mammalian cells. Proteins are detected with human-specific ORF1p polyclonal antibodies. GAPDH used as loading control. Molecular weights in kilodaltons (kDa) are shown on the right. Red asterisk is an ORF1p heterodimer.

Figure S3.7

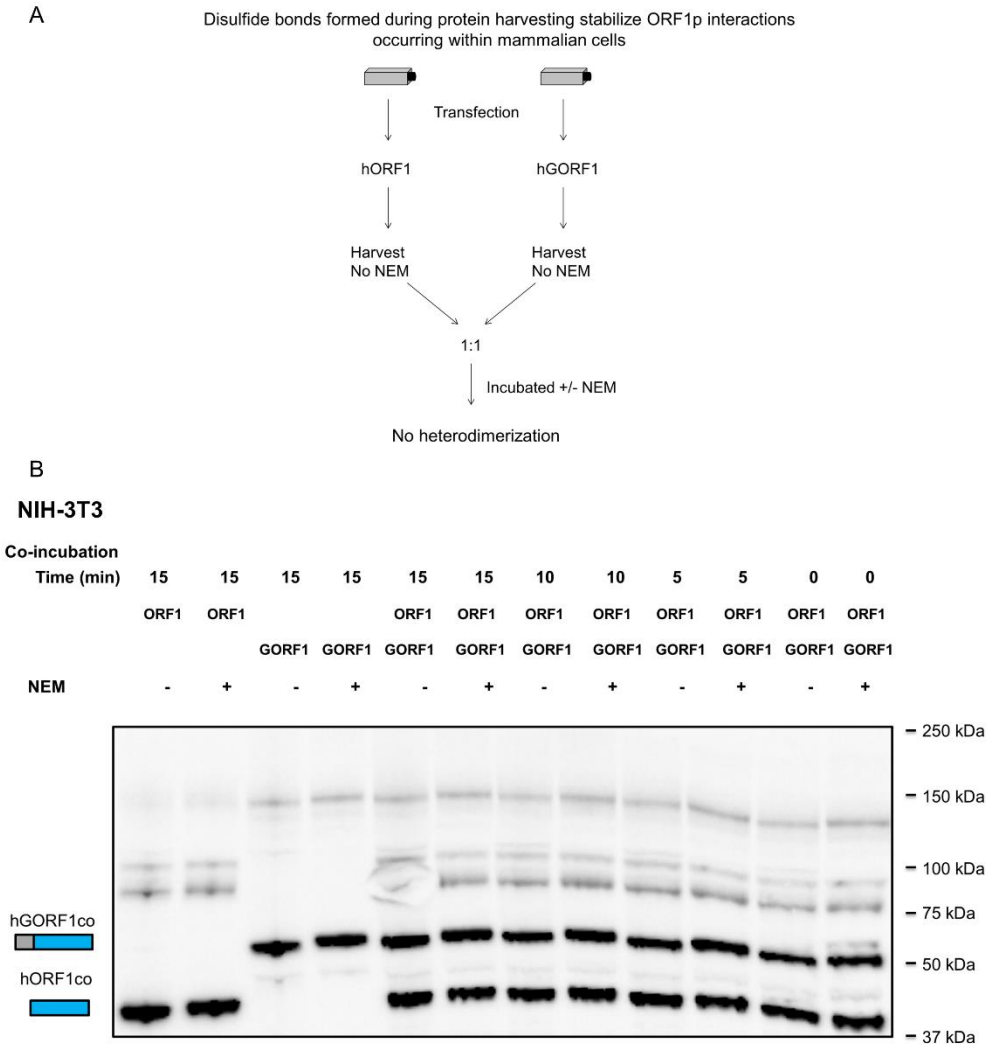

Figure S3.7. Analysis of ORF1p species as a function of incubation time. A. Schematic of experiment: Cells are transfected individually with either hORF1co or hGORF1co expression plasmids and processed separately. Samples are combined prior to loading on western blot for indicated co-incubation time (see B), with or without NEM (25mM). **B.** Western blot analysis of ORF1p species formed during protein harvesting detects monomer, dimers, and trimers. No heterodimer formation was observed under these experimental conditions. Proteins are detected with human-specific ORF1p polyclonal antibodies. GAPDH used as loading control. Molecular weights in kilodaltons (kDa) are displayed on the right.

Figure S3.8

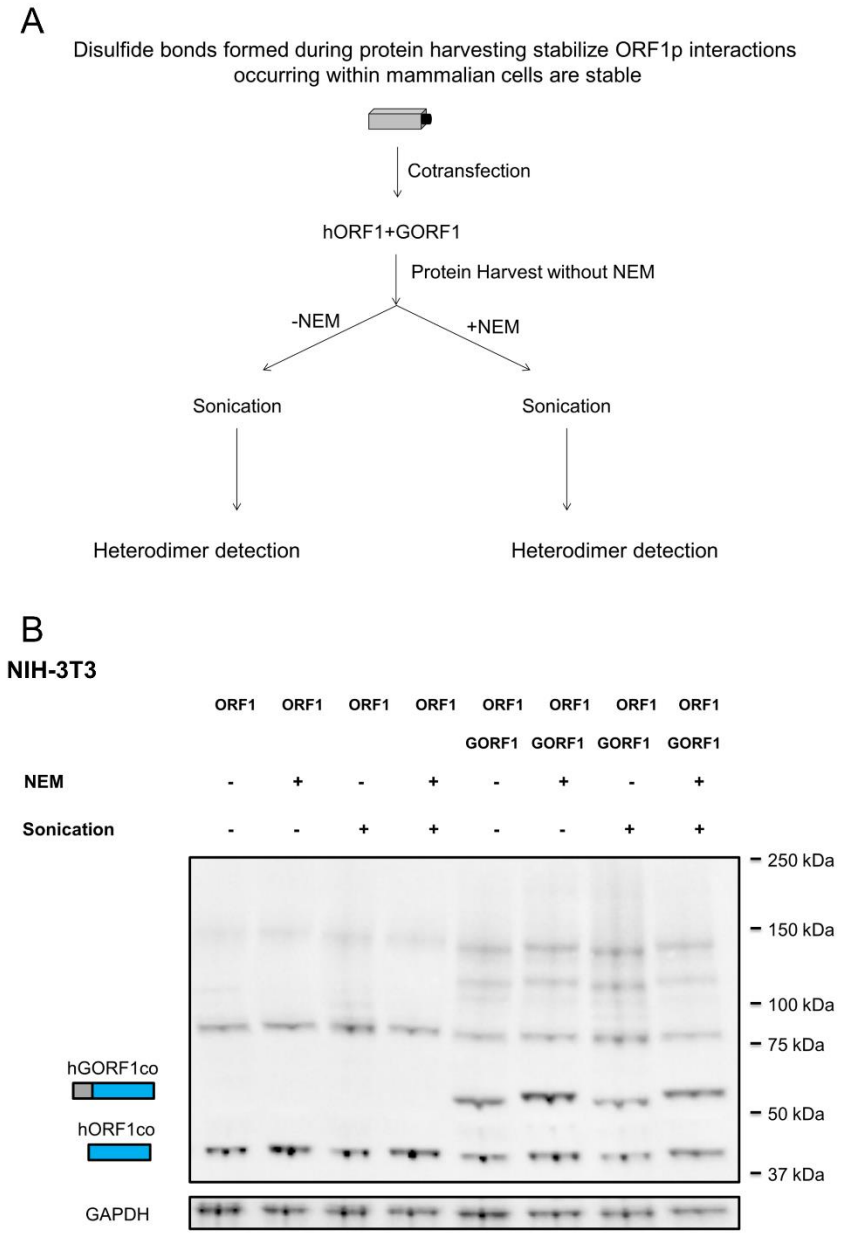

Figure S3.8. Sonication following protein harvesting does not effect observed ORF1p species. A. Schematic of experiment: cells are transfected individually or cotransfected with hORF1co and/or hGORF1co, and lysates are harvested without NEM. After processing, samples are split into two tubes, with one tube having 25 mM NEM added and the other with an equivalent amount of ethanol. Samples were then sonicated and fractionated on an SDS-PAGE gel under non-reducing conditions (see materials and methods). **B.** Western blot analysis of ORF1p species formed during protein harvesting detects heterodimers which are stable regardless of sonication or the presence/absence of NEM. Proteins are detected with human-specific ORF1p polyclonal antibodies. GAPDH used as loading control. Molecular weights in kilodaltons (kDa) are displayed on the right.

Table S3.1

	Abbreviation	Full name
Human		
	hORF1co	Human codon-optimized ORF1
	hGORF1co	GAL4 tagged-human codon-optimized ORF1
	hVORF1co	VP16 tagged-human codon-optimized ORF1
	hORF1wt	Human L1.3 wild-type ORF1
	hGORF1wt	GAL4 tagged-human codon-optimized ORF1
	hVORF1wt	VP16 tagged-human codon-optimized ORF1
	hL1wt	Full length L1.3 construct
Mouse		
	mL1wt	Full length mouse L1 Spa construct
	mORF1co	Mouse codon-optimized ORF1 Spa
	mGORF1co	GAL4 tagged-mouse codon-optimized ORF1 Spa
	mVORF1co	VP16 tagged-mouse codon-optimized ORF1 Spa
	mGORF1wt	GAL4 tagged-mouse L1 Spa ORF1 wild-type
	mVORF1wt	VP16 tagged-mouse L1 Spa ORF1 wild-type

See Material and methods for additional information

Table 3.1. Constructs utilized in the study of ORF1p self-interaction. Constructs expressing ORF1p of human and mouse origin are shown. Abbreviations are shown on the left with the parent construct given on the right.

APPENDIX 3: Information Supplemental to "Chapter 4: Truncated ORF1 proteins can suppress LINE-1 retrotransposition in *trans*"

Figure S4.1

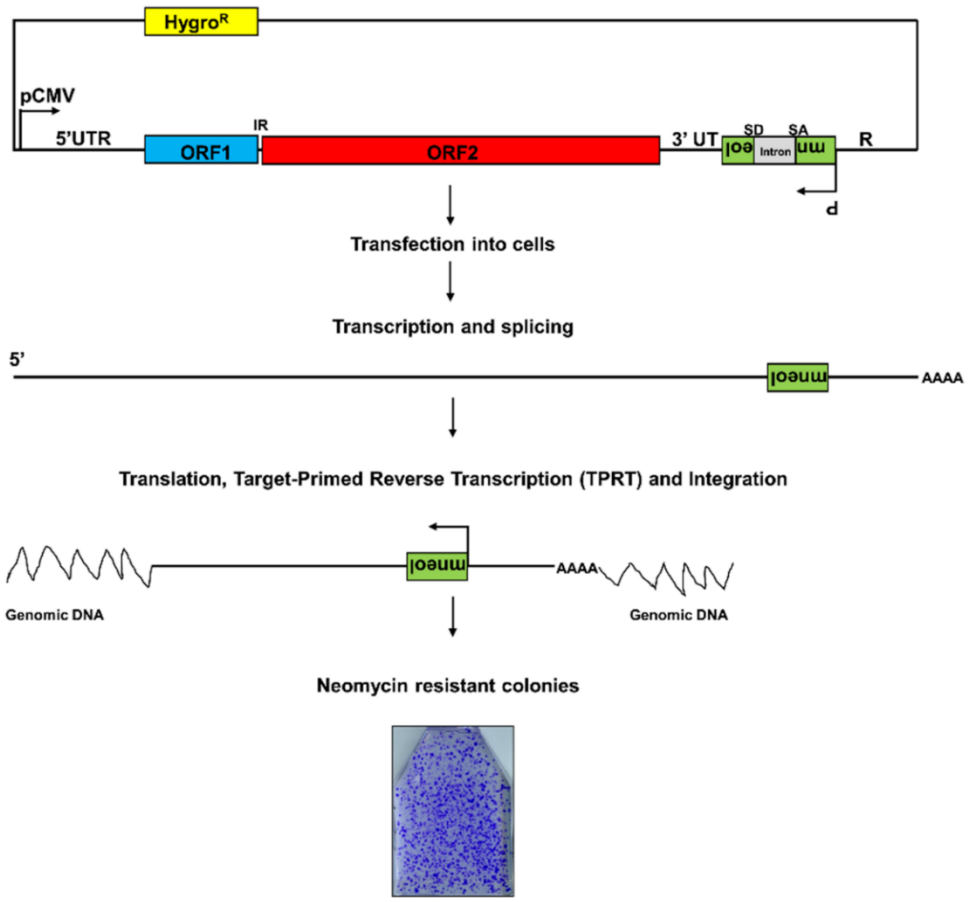

Figure S4.1. L1 retrotransposition assay A. (top) Schematic of a L1Neo plasmid (*7*). pCMV is a CMV promoter, 5' UTR is the 5' untranslated region of the L1.3 element, ORF1 is Open Reading Frame 1, IR is an intergenic region, ORF2 is Open Reading Frame 2, 3'UTR is the 3' untranslated region of the L1.3 element, This 3'UTR contains a cassette that includes a promoter, a neomycin resistance gene (mneol) disrupted by an intron (SD=splice donor and SA=splice acceptor) and a polyA site. The promoter, neo gene, and pA site are in the reverse orientation relative to L1 transcription. The intron interrupting the neo gene is in the same orientation as the L1 transcript and therefore it is removed from the mature L1 mRNA. The plasmid also contains hygromycin resistance gene that is used to generate HeLa cells stably expressing various L1 elements. When this plasmid is transfected into cells, the L1 is transcribed and the intron is spliced. The L1 mRNA is translated and the resulting ORF1 and ORF2 proteins associate with their parental mRNA to form a ribonucleoprotein particle (RNP), which gains access to the nucleus. This L1 RNP undergoes Target-Primed Reverse Transcription (TPRT) and integration into the genome. The neomycin resistance gene is now able to be expressed to provide neomycin resistance to the cell leading to the formation of a Neo-resistant colony in the presence of neomycin (G418) selection.

Figure S4.2

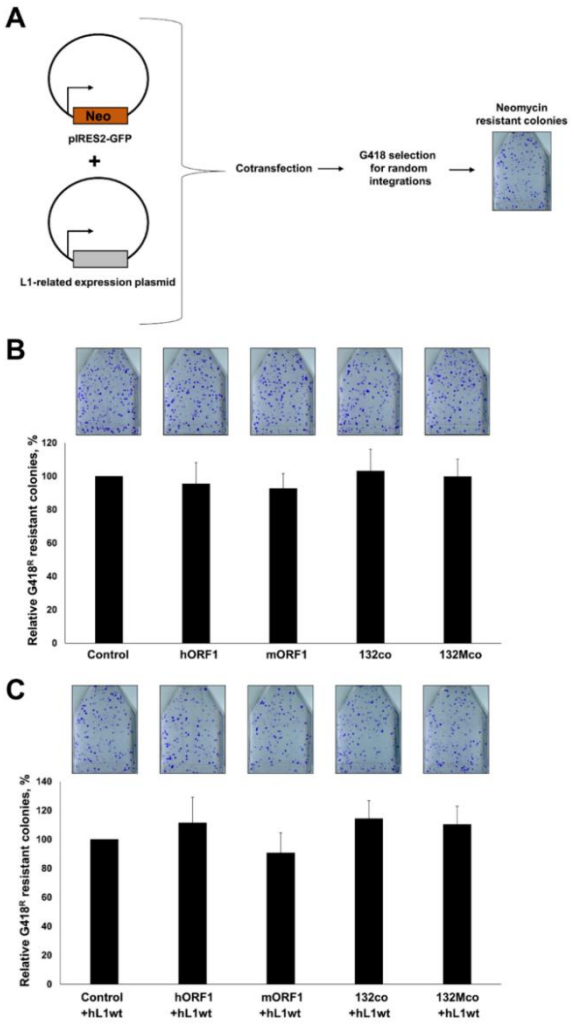

Figure S4.2. Toxicity assay in HeLa cells using ORF1 plasmids or as a cotransfection using ORF1 plasmids along with a plasmid expressing wild-type human L1. A. A schematic of the experimental approach as was previously reported (*1*, *2*). pIRES2-GFP plasmid contains a neomycin resistance (Neo) gene. This plasmid can be cotransfected with various L1-related expression plasmids to determine whether they have any negative effect on cell viability. **B.** Results of toxicity assay using a pIRES construct containing neomycin resistance cotransfected with empty (control), hORF1, mORF1, 132co or 132Mco expression plasmids. The number of G418 resistant colonies was normalized to the control flask (pIRES construct cotransfected with an empty plasmid) for each independent experiment. No statistical significance was detected between the control and experimental constructs (n=3, t-test, $p<0.05$). **C.** Results of toxicity assay in HeLa cells using a pIRES construct containing neomycin resistance cotransfected with a plasmid containing wild-type human L1 and either empty(control), hORF1, mORF1, 132co or 132Mco expression plasmids. The number of G418 resistant colonies was normalized to the control flask (pIRES construct cotransfected with the empty plasmid) for each independent experiment. No statistical significance was detected between the control and experimental constructs (n=3, t-test, $p<0.05$).

Figure S4.3

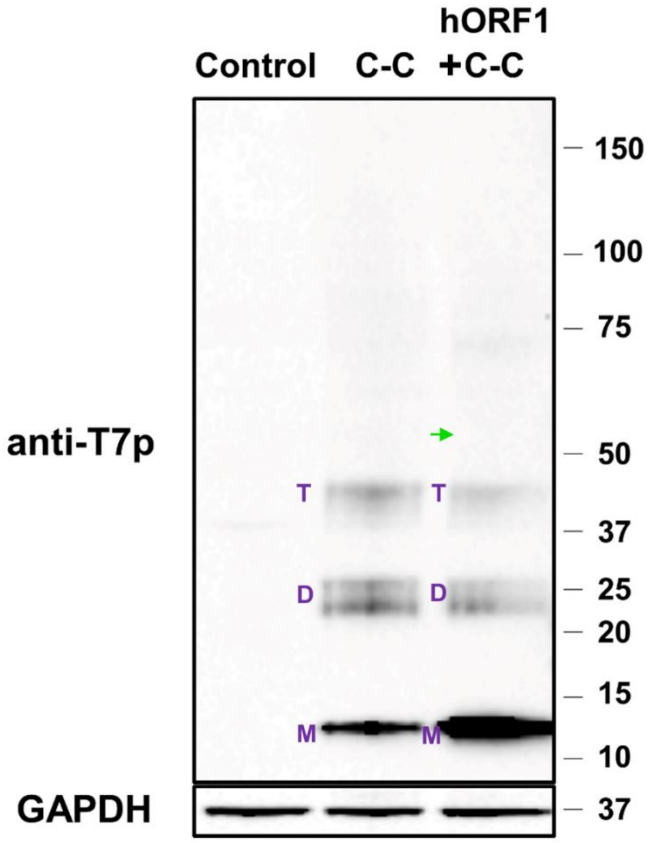

Figure S4.3. Human C-C protein does not form heterodimers with the full-length human ORF1p when coexpressed in HeLa cells. Western blot analysis (non-reducing conditions, anti-T7 tag monoclonal antibodies (anti-T7p)) of the T7-tagged C-C protein generated in HeLa cells transfected with either the C-C expression plasmid alone or the C-C and the hORF1 expression plasmids. Expected heterodimer product would have a molecular weight of 53.9 kDa (green arrow). Purple letters denote monomers "M", dimers "D" or trimers "T" formed by the C-C ORF1p. GAPDH detection was used as a loading control. Numbers on the right of the image are molecular weight markers in kDa.

Figure S4.4

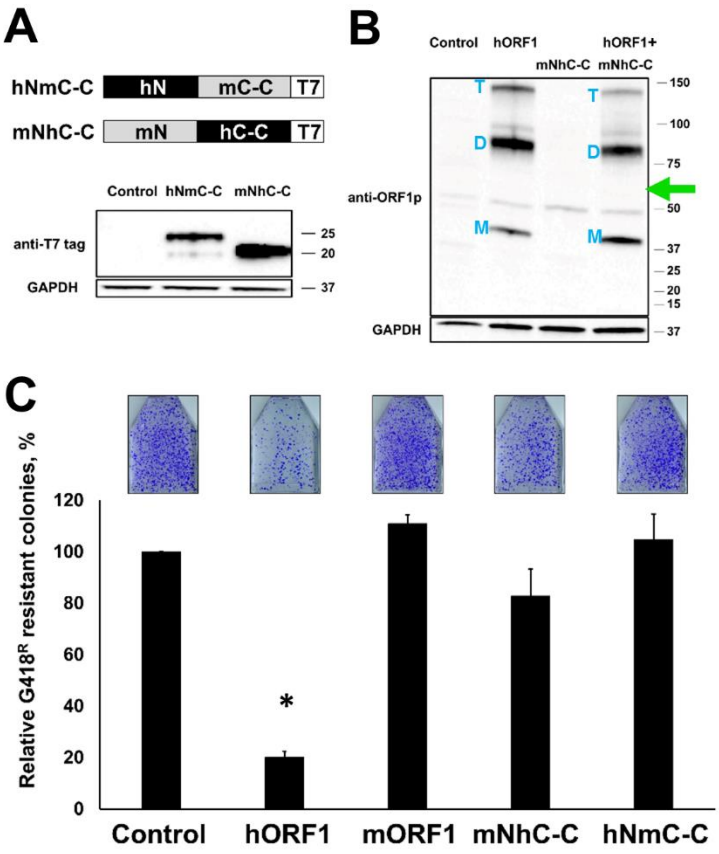

Figure S4.4. Analysis of human/mouse ORF1p chimeric proteins expression and their effect on human L1 retrotransposition. (A) (top) Schematic of human/mouse ORF1 chimeric constructs. Human N-terminus and mouse ORF1 coiled-coil domain (hNmC-C) as well as mouse N-terminus and human ORF1 coiled-coil domain (mNhC-C) constructs are shown. T7 indicates a T7 tag. **(bottom)** Western blot analysis (reducing conditions) using T7 tag specific antibodies of proteins generated from codon-optimized chimeric ORF1 constructs, mNhC-C and hNmC-C, transiently transfected in HeLa cells. GAPDH detection was used as a loading control. Molecular weight markers are shown on the right in kDa. Control lane indicates cells transfected with an empty plasmid. **(B)** Western blot analysis (non-reducing conditions) of proteins generated from codon-optimized hORF1 and mNhC-C constructs transiently transfected in HeLa cells. anti-ORF1p polyclonal antibodies (anti-ORF1p) were used for detection. Molecular weight markers are shown on the right in kDa. GAPDH was used as a loading control. The position of expected heterodimer product is indicated by the green arrow (58.5 kDa). Blue letters denote monomers "M", dimers "D" or trimers "T" of full-length ORF1p generated from hORF1 plasmid. **(C)** Results of the L1 retrotransposition assay in HeLa cells cotransfected with the L1Neo plasmid and either empty (Control), hORF1, mORF1, mNhC-C or hNmC-C expression plasmids. The total G418 resistant colony counts were normalized to the Control flask for each independent experiment (L1 construct along with empty plasmid supplied in *trans*). Asterisk (*) denotes statistical significance between listed constructs and the control (n=3, t-test, $p<0.05$). For mNhC-C, the p value is 0.052.

Figure S4.5

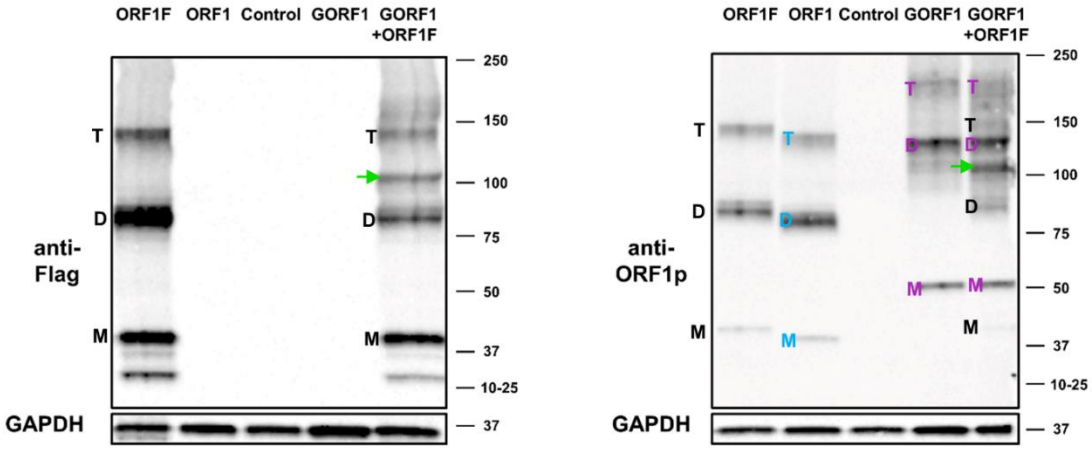

Figure S4.5. Flag-tagged ORF1p form heterodimers with Gal4-fused ORF1p. (left) Western blot analysis (non-reducing conditions, Flag-tag specific antibodies (anti-Flag)) of Flag-tagged ORF1p, GAL4-fused ORF1p and ORF1p expressed in HeLa cells from their respective plasmids. Green arrow denotes expected molecular weight of a heterodimer between the Flag-tagged and GAL4-fused ORF1p (100 kDa). Control lane are cells transfected with empty plasmid. Molecular weight markers are shown on the right in kDa. GAPDH is used as a loading control. M, D, T letters mark monomers, dimers, and trimers generated by the ORF1F protein. **(right)** Western blot analysis of the same samples shown in (A) using anti-ORF1p antibodies. Green arrow denotes expected molecular weight of a heterodimer between a Flag-tagged and a GAL4-fused ORF1p. Control lane contains lysates extracted from cells transfected with an empty plasmid. Molecular weight markers are shown on the right in kDa. GAPDH is used as a loading control. M, D, and T letters mark monomer, dimer and trimer species for each protein.

Figure S4.6

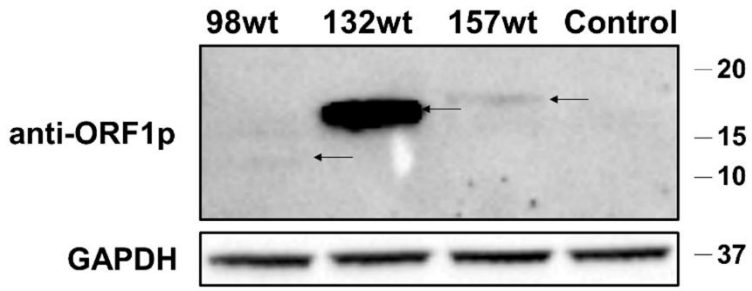

Figure S4.6. Expression of truncated ORF1 proteins generated from plasmids containing wild-type (non-codon-optimized) L1 sequence. High exposure of a Western blot of truncated ORF1 protein generated from 98, 132, and 157 constructs containing wild-type (wt) L1 sequence transiently transfected in HeLa cells (shorter exposure of the western shown in Figure 5). Control lane indicates cells transfected with an empty plasmid. GAPDH is used as a loading control. Positions of molecular weight markers are indicated on the right in kDa. Anti-ORF1p polyclonal antibodies were used for western blot analysis. Black arrows denote expected molecular weight of truncated ORF1 proteins generated from their respective expression plasmids (98 (11.3 kDa), 132 (15.5 kDa), 157 (18.7 kDa)).

Figure S4.7

A

0.2 conditions

Construct	Amount (µg)	Construct	Amount (µg)	Construct	Amount (µg)
hL1wt Neo	0.4	98wt	0.2	pBud	0.2
hL1wt Neo	0.4	157wt	0.2	pBud	0.2
hL1wt Neo	0.4	98wt	0.2	157wt	0.2
hL1wt Neo	0.4	pBud	0.2	pBud	0.2
hL1wt Neo	0.4	hORF1co	0.2	pBud	0.2
hL1wt Neo	0.4	mORF1co	0.2	pBud	0.2
hL1wt Neo	0.4	98co	0.2	pBud	0.2
hL1wt Neo	0.4	157co	0.2	pBud	0.2
hL1wt Neo	0.4	98co	0.2	157co	0.2

0.4 conditions

Construct	Amount (µg)	Construct	Amount (µg)	Construct	Amount (µg)
hL1wt Neo	0.4	98wt	0.4	pBud	0.4
hL1wt Neo	0.4	157wt	0.4	pBud	0.4
hL1wt Neo	0.4	98wt	0.4	157wt	0.4
hL1wt Neo	0.4	pBud	0.4	pBud	0.4
hL1wt Neo	0.4	hORF1co	0.4	pBud	0.4
hL1wt Neo	0.4	mORF1co	0.4	pBud	0.4
hL1wt Neo	0.4	98co	0.4	pBud	0.4
hL1wt Neo	0.4	157co	0.4	pBud	0.4
hL1wt Neo	0.4	98co	0.4	157co	0.4

Figures 5 and 3

Construct	Amount (µg)	Construct	Amount (µg)
hL1wt Neo	0.4	98wt	0.4
hL1wt Neo	0.4	157wt	0.4
hL1wt Neo	0.4	98wt	0.4
hL1wt Neo	0.4	pBud	0.4
hL1wt Neo	0.4	hORF1co	0.4
hL1wt Neo	0.4	mORF1co	0.4
hL1wt Neo	0.4	98co	0.4
hL1wt Neo	0.4	157co	0.4
hL1wt Neo	0.4	98co	0.4

B

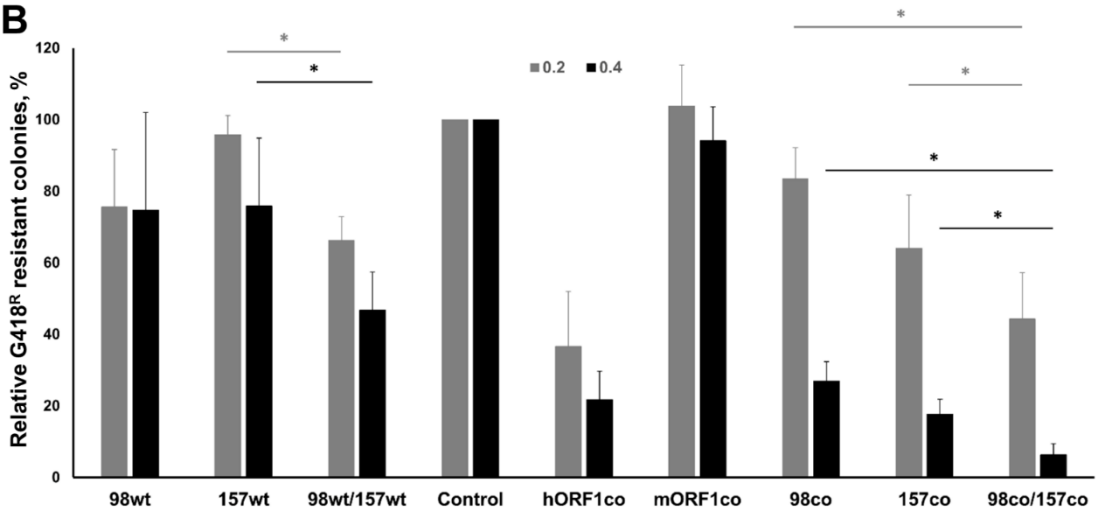

Figure S4.7. The effect of truncated ORF1p co-expression on L1 retrotransposition. (A) Experimental approach to comparing the effect of individual and combined expression of truncated 97 and 156 proteins on L1 retrotransposition. Wt and co indicate plasmids containing wild-type non-codon-optimized or codon-optimized L1 sequence, respectively. DNA amounts (µg) used for transfection are indicated for each plasmid. pBud is an empty plasmid used as filler. **(B)** L1 retrotransposition in HeLa cells using a Neo-tagged human full-length wild-type L1 expression plasmid cotransfected with an empty (control), mORF1, hORF1, 98wt and/or 157wt, or 98co and/or 157co expression plasmids. Total number of G418 resistant colonies was normalized to the control flask (L1 construct cotransfected with empty plasmid) for each independent experiment. Asterisk (*) denotes statistical significance between listed constructs done under the same conditions (n=3, t-test, $p<0.05$).

Figure S4.8

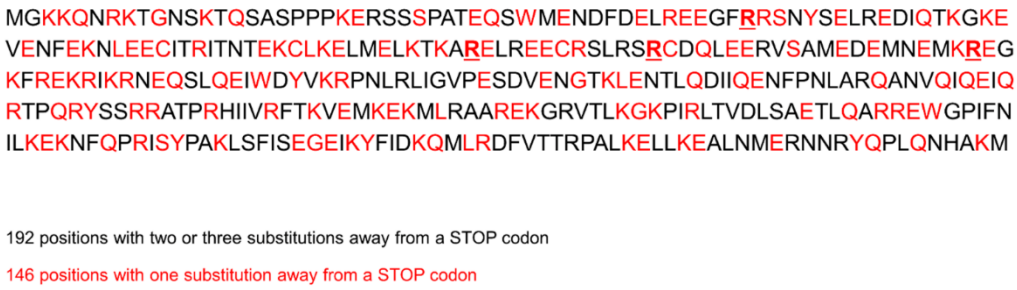

Figure S4.8. L1.3 ORF1 sequence contains codons that are one or two substitutions away from becoming a stop codon. L1.3 ORF1 sequence was analyzed for their potential to generate a stop codon. Amino acids listed in red are encoded by codons that are one substitution away from becoming a stop codon (approximately 43%). The amino acid positions that are bolded and underlined are those that are the most common stop positions shown in Figure 6. Amino acids listed in black are encoded by codons that are two or three substitutions away from generating a stop codon (approximately 57%).

Figure S4.9

chromosome	position	position	orient.	AA Position	Subfamily	Allelic Frequency
chr5	89958428	89964454	-	110	PA3	0.37
chr5	65552645	65558835	+	130	PA3	0.10
chr6	133463213	133469243	-	130	PA3	0.10
chr5	25113847	25119863	-	110	PA2	0.58

Figure S4.9. The presence of a stop codon within the ORF1 sequence of fixed human L1 loci can be polymorphic. Manual analysis of SNPs in the ORF1 sequence of 50 full-length L1 loci (reference build hg19 of the UCSC genome browser) identified 4 L1 loci in which the presence of stop codons in ORF1 between amino acids 110-130 is polymorphic.

Chromosome is Chromosome #

Position indicates the coordinates of each L1 locus in the human genome based on the hg19 build in the UCSC genome browser

Orient. The relative orientation of the L1 loci, sense (+) or antisense (-) is listed

AA Position indicates the position of the stop codon relative to the ORF1 protein sequence

Subfamily indicates the L1 subfamily

Allelic frequency indicates the allelic frequency of the polymorphism resulting in a stop codon as listed in the reference build hg19 of the UCSC genome browser

Figure S4.10

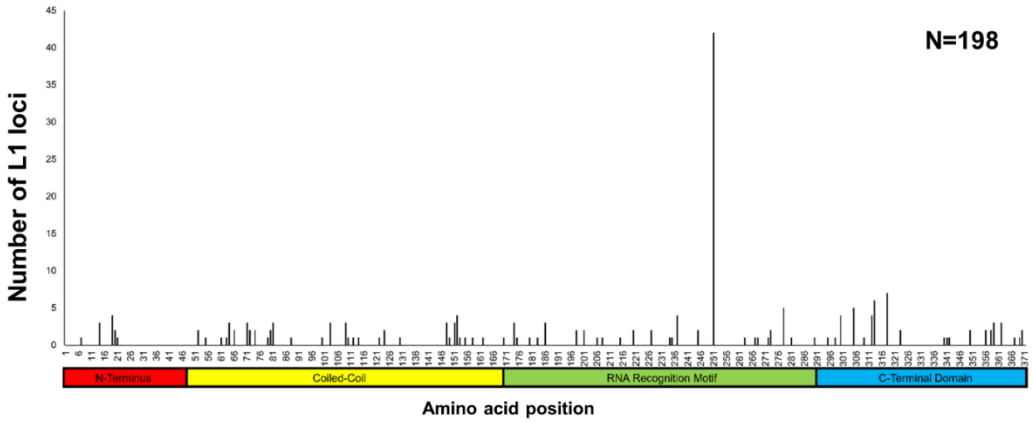

Figure S4.10. Bioinformatic analysis of stop codon positions in retrotranspositionally-incompetent mouse L1 loci. (A) (top) Full-length mouse L1 loci identified using the L1Base database (*6*) were analyzed to determine the position of the first stop codon in their ORF1 sequence. The amino acid position 251 is the most common stop position as determined by analysis of 198 L1 loci (n=198). **(bottom)** A schematic of the mouse ORF1 protein is shown below the graph.

Figure S4.11

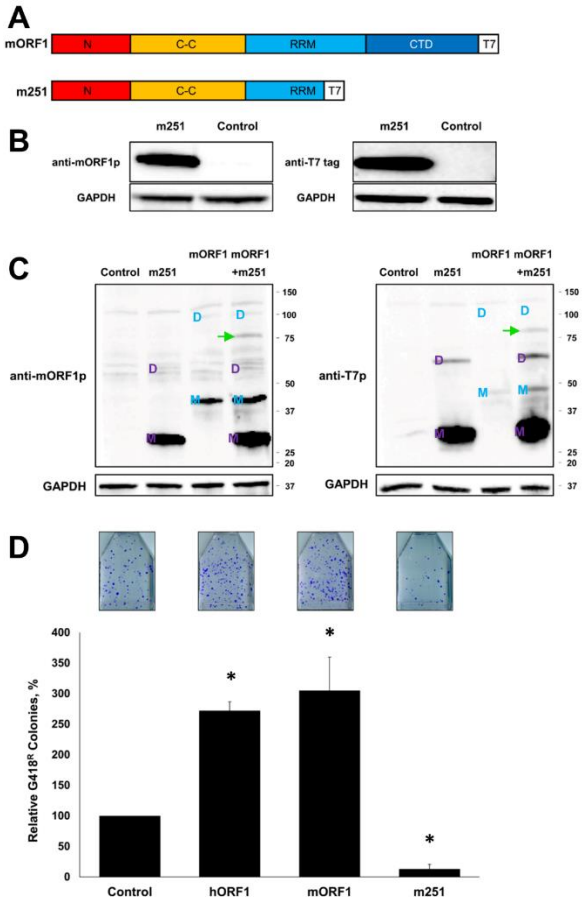

Figure S4.11. Analysis of truncated mORF1p expression and its effect on mouse L1 retrotransposition. (A) A schematic of the full-length mouse ORF1 (mORF1) and C-terminally truncated mouse ORF1 (m251) constructs (codon-optimized). **(B)** Western blot analysis of the total lysates collected using reducing conditions from HeLa cells transiently transfected with codon-optimized m251 construct. m251 protein is detected using anti-mouse ORF1p or anti-T7 polyclonal antibodies (anti-mORF1p, anti-T7 tag). Control lane indicates cells transfected with an empty plasmid. GAPDH was used as a loading control. **(C)** Western blot analysis of proteins harvested under non-reducing conditions from HeLa transfected with the codon-optimized m251 and/or mORF1 constructs. Green arrow denotes the position of the expected m251/mORF1p heterodimer (expected molecular weight is 77 kDa). m251 dimer (D) and monomer (M) are listed in purple (30 and 60 kDa respectively) and mORF1 dimer (D) and monomer (M) are listed in blue (47kDa and 94 kDa respectively). Control lane indicates cells transfected with an empty plasmid. Anti-mouse ORF1p (anti-mORF1p) or anti-T7 (anti-T7 tag) polyclonal antibodies were used to detect full-length and truncated (T7 tagged) mouse ORF1 proteins. Molecular weight markers are shown on the right in kDa. **(D)** Result of L1 retrotransposition assay using a Neo tagged mouse wild-type non-codon-optimized full-length L1 plasmid cotransfected with ab empty (control), hORF1, m251 or mORF1 expression plasmids in HeLa cells. Total number of G418 resistant colonies detected for each experimental flask was normalized to the number of G418 resistant colonies identified on the control flask (mouse L1 construct cotransfected with empty plasmid) for each independent experiment. Asterisk (*) denotes statistical significance between listed constructs and the control (n=3, t-test, $p<0.05$).

Figure S4.12

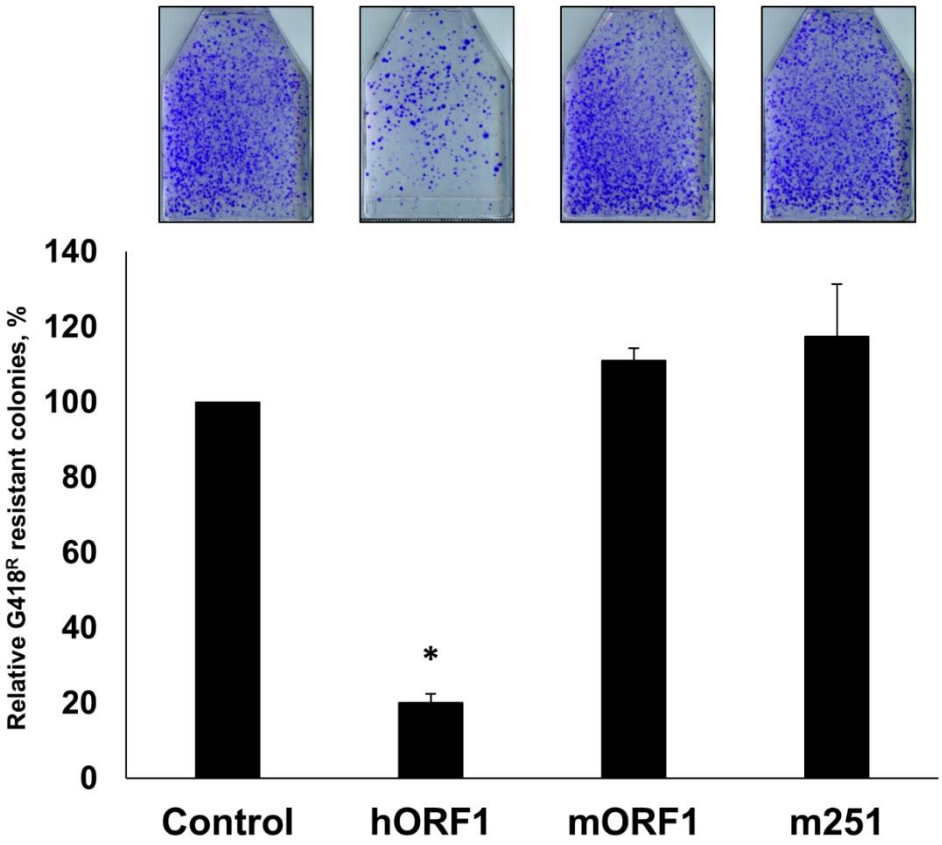

Figure S4.12. Truncated mORF1 251 protein does not suppress human L1 retrotransposition. Results of L1 retrotransposition in HeLa cells using a Neo-tagged human full-length wild-type L1 expression plasmid cotransfected with an empty (control), hORF1, mORF1, or m251 expression plasmids. Total number of G418 resistant colonies detected for each experimental flask was normalized to the number of G418 resistant colonies identified on the control flask (human L1 construct cotransfected with empty plasmid) for each independent experiment. Asterisk (*) denotes statistical significance between listed constructs and the control (n=3, t-test, $p<0.05$).

Figure S4.13

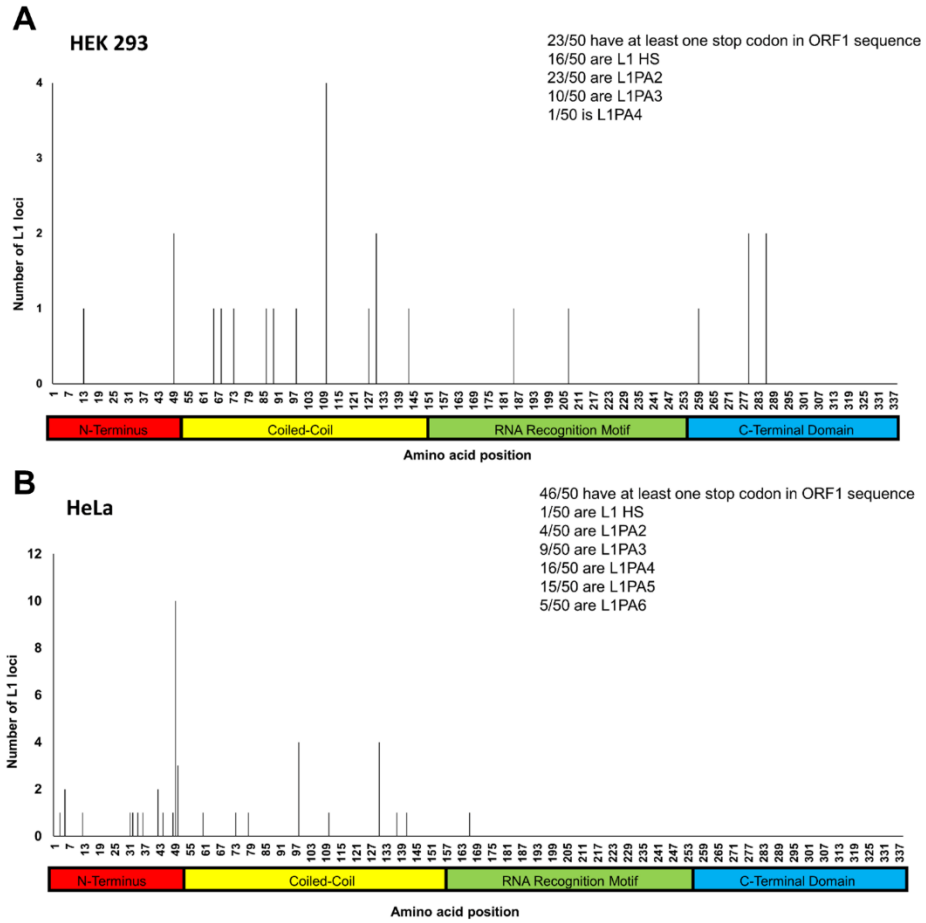

Figure S4.13. Endogenous L1 loci containing at least one stop codon within the ORF1 sequence are expressed in HEK 293 and HeLa cells. (A) Sequences of L1 loci determined by RNAseq (*8*) to be expressed in HEK 293 cells (Supplementary Table 1) were used to determine whether they contain stop codons in their ORF1 sequence. The first stop codon position identified in the ORF1 sequence of these L1 loci was plotted. A schematic of the human ORF1 protein domains is shown below the graph. The breakdown of L1 subfamilies of the top 50 expressed L1 loci is shown. **(B)** The same analysis as in (A) performed using the RNA sequencing data from HeLa cells (Supplementary Table 2).

Figure S4.14

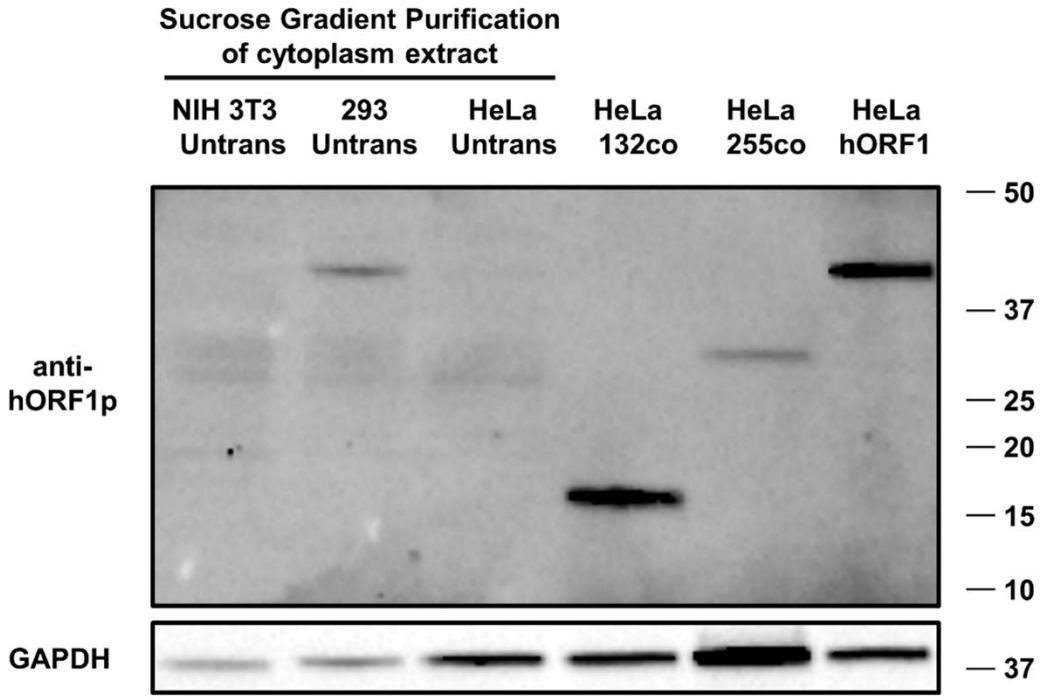

Figure S4.14. Analysis of endogenous L1 ORF1p expression in mammalian cell lines. Sucrose gradient purification of polyribosome-rich fraction from NIH 3T3, HEK 293 and HeLa cells (see Supplementary Materials and Methods) was analyzed by western blot with anti-human ORF1 antibodies (*9*). ORF1 proteins generated from codon-optimized 132, 255 and hORF1 constructs transiently transfected in HeLa cells were used as molecular markers. anti-ORF1p polyclonal antibodies (anti-ORF1p) detect a band consistent with the expected molecular weight of a full-length ORF1 in HEK 293 cells. Numbers listed on the right indicate positions of molecular weight markers in kDa. GAPDH was used as a loading control.

Figure S4.15

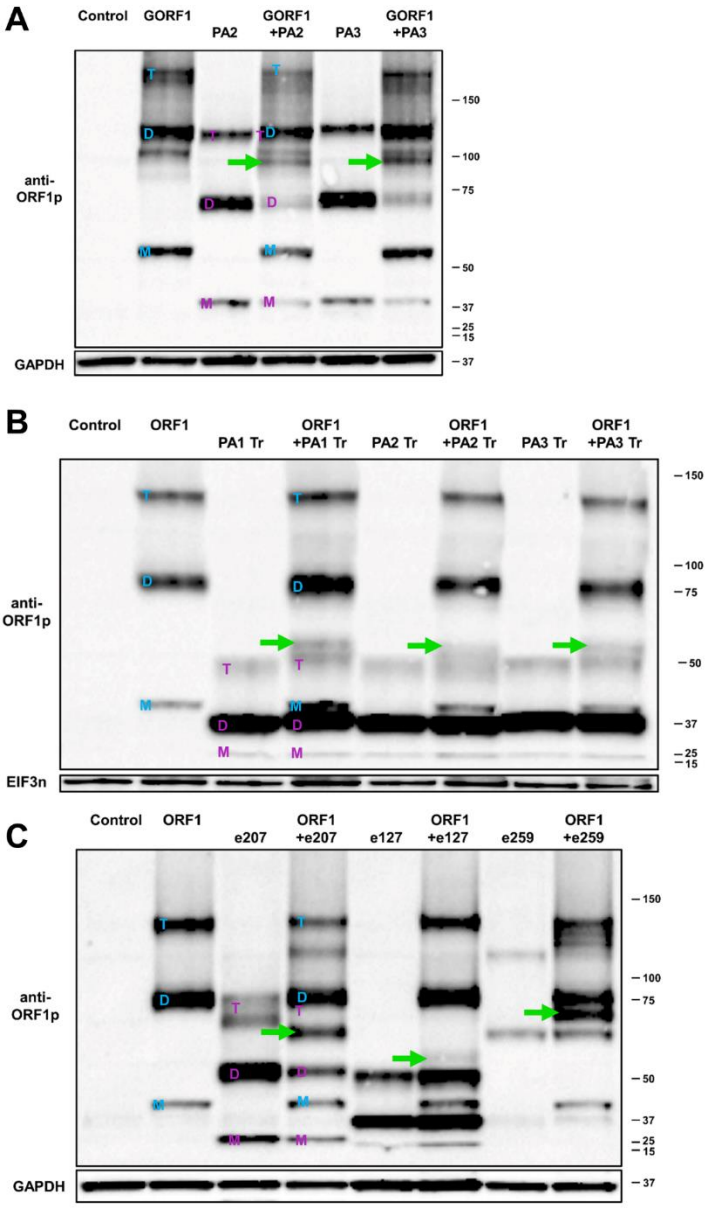
A
Control
GORF1
PA2
GORF1 +PA2
PA3
GORF1 +PA3
anti-ORF1p
GAPDH
150
100
75
50
37
25
15
B
Control
ORF1
PA1 Tr
ORF1 +PA1 Tr
PA2 Tr
ORF1 +PA2 Tr
PA3 Tr
ORF1 +PA3 Tr
anti-ORF1p
EIF3n
C
Control
ORF1
e207
ORF1 +e207
e127
ORF1 +e127
e259
ORF1 +e259
anti-ORF1p
GAPDH

Figure S4.15. Analysis of expression of truncated ORF1 proteins representing older L1PA subfamilies and their potential to suppress L1 retrotransposition. (A) Western blot analysis (non-reducing conditions) of heterodimerization of proteins generated from codon-optimized GAL4-fused ORF1 (GORF1), L1PA2 ORF1 (PA2) and L1PA3 ORF1 (PA3) constructs transiently cotransfected in HeLa cells. Green arrows denote the position of expected heterodimers: (97 kDa). Control lane indicates cells transfected with an empty plasmid. Western blot analysis was performed using anti-ORF1p polyclonal antibodies (anti-ORF1p). Numbers listed on the right indicate positions of molecular weight markers in kDa. GAPDH was used as a loading control. Blue letters denote GAL4-fused ORF1p monomers "M", dimers "D" or trimers "T". Purple letters denote L1PA2 protein monomers "M", dimers "D" or trimers "T". **(B)** Western blot analysis (non-reducing conditions) of heterodimerization of proteins generated from codon-optimized ORF1 (ORF1), L1PA1 130 (PA1 Tr), L1PA2 130 (PA2 Tr) and L1PA3 130 (PA3 Tr) constructs transiently cotransfected in HeLa cells. Green arrows denote the position of expected heterodimers: (55 kDa). Control lane indicates cells transfected with an empty plasmid. Western blot analysis was performed using anti-ORF1p polyclonal antibodies (anti-ORF1p). Numbers listed on the right indicate positions of molecular weight markers in kDa. GAPDH was used as a loading control. Blue letters denote ORF1p monomers "M", dimers "D" or trimers "T". Purple letters denote truncated PA1 ORF1 monomers "M", dimers "D" or trimers "T". **(C)** Western blot analysis (non-reducing conditions) of heterodimerization of proteins generated from codon-optimized ORF1 (ORF1) or endogenous L1 ORF1 constructs (e207, e127 and e259) transiently cotransfected in HeLa cells. Green arrows denote the position of expected heterodimers: for e207 (64.4 kDa), e127 (54.8 kDa) and e259 (70.2 kDa). Control lane indicates cells transfected with an empty plasmid. Western blot analysis was performed using anti-ORF1p polyclonal antibodies (anti-ORF1p). Numbers listed on the right indicate positions of molecular weight markers in kDa. GAPDH was used as a loading control. Blue letters denote ORF1p monomers "M", dimers "D" or trimers "T". Purple letters denote truncated e207 monomers "M", dimers "D" or trimers "T".

Figure S4.16

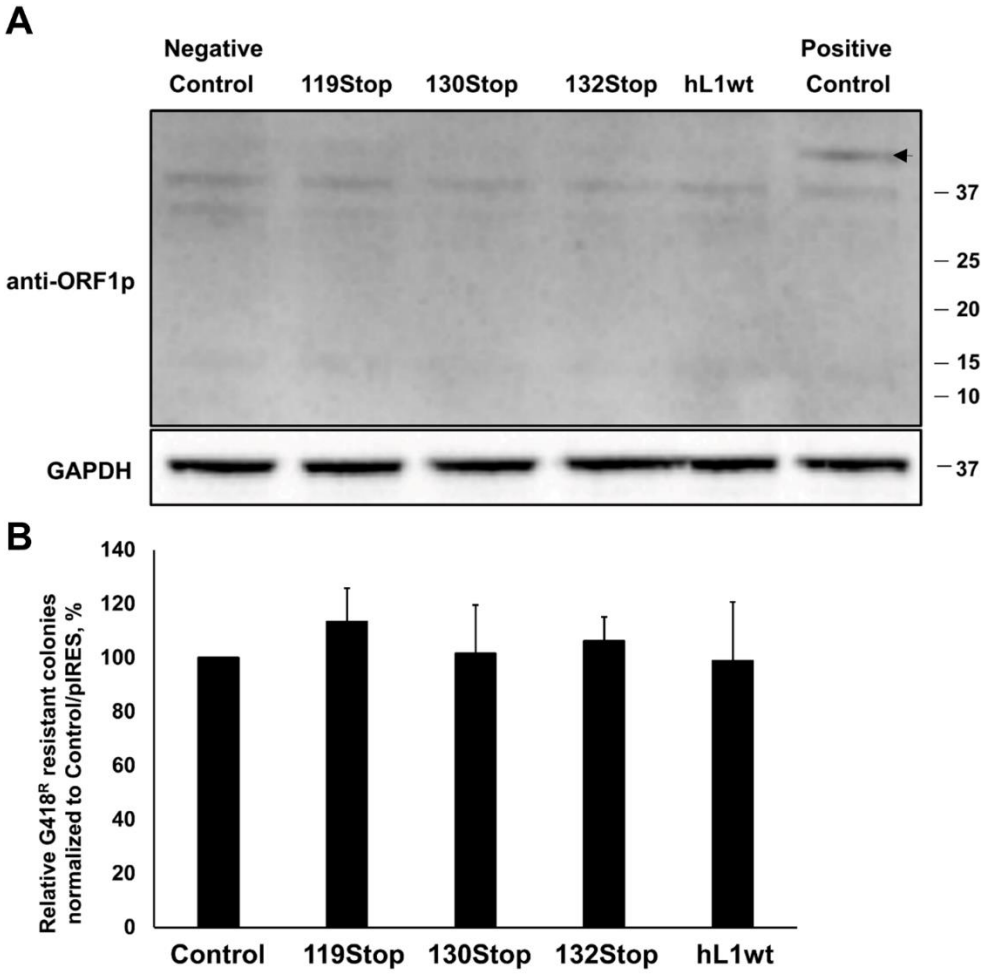

Figure S4.16. Loss of ORF1p expression in engineered HeLa cells harboring L1Stop elements results in the loss of suppression of L1 retrotransposition. (A) Western blot analysis of ORF1 proteins in HeLa cells harboring full-length L1 (hL1wt), 119Stop, 130Stop, 132Stop, or an empty plasmid (Control) after several months of propagation. Positive control (black arrow) of ORF1 protein generated from a HeLa cell line stably expressing a full-length L1. Anti-ORF1p polyclonal antibodies were used. GAPDH was used as a loading control. Molecular weight markers are shown on the right in kDa. **(B)** Results of wild-type human Neo-tagged L1 retrotransposition in HeLa cell lines described above. Total number of G418 resistant colonies detected on experimental flasks was normalized to the number of colonies on the Control flask for each independent experiment. No statistical significance was detected between the control and experimental cell lines.

Table S4.1

CHR	Left	Right	Strand	Stop	Position	Subfamily
chr22	29059274	29065303	+	No		HS
chr13	37724090	37730119	-	Yes	259	PA2
chr7	65751841	65757868	-	No		HS
chr1	174812365	174818381	-	Yes	207	PA2
chr2	118894577	118900596	-	Yes	89	PA2
chr15	64618860	64624899	+	Yes	286	PA2
chr2	71638605	71644631	-	No		HS
chr13	39574955	39581012	-	No		PA2
chr7	49719865	49725896	-	No		HS
chr12	76713405	76719506	+	Yes	49	PA3
chr2	197770314	197776343	+	No		HS
chrX	154745678	154751709	-	No		HS
chr11	24349498	24355547	+	No		HS
chr5	89958428	89964454	-	Yes	110	PA3
chr7	76357803	76363830	+	Yes	73	PA3
chr4	21161014	21167044	+	No		HS
chr2	66618263	66624282	+	Yes	185	PA2
chr3	51010569	51016603	-	No		PA3
chr9	89656798	89662815	+	No		PA3
chr20	23406746	23412777	+	No		HS
chr9	85664455	85670484	-	No		HS
chr12	96709723	96715749	+	Yes	127	PA2
chr20	8575749	8581774	-	No		PA2
chr2	175269187	175275210	-	No		PA2
chr5	172829800	172835828	-	No		HS
chr17	66404265	66410283	-	Yes	98	PA3
chr1	43580199	43586164	-	Yes	68	PA2
chr13	50754432	50760463	-	No	NSC	PA2
chr7	113416178	113422207	-	No		HS
chr15	51323827	51329846	-	No		PA3
chr10	36147597	36153623	+	No		PA3
chr7	96475963	96481992	-	No		HS
chr5	108595091	108601119	+	No		HS
chrX	11953208	11959431	-	No		HS
chr6	2418009	2424034	-	No		HS
chrX	130092906	130098931	+	No		PA2
chr14	55337083	55343223	+	Yes	NSC	PA4
chr16	72661174	72667197	+	Yes	286	PA2
chr18	39211707	39217733	+	Yes	143	PA2

chr9	17285002	17291032	-	Yes	13	PA2
chr2	135408125	135414144	-	Yes	110	PA2
chr4	147271171	147277203	+	Yes	49	PA2
chr5	74991524	74997580	+	Yes	65	PA3
chr2	198552257	198558285	-	Yes	130	PA2
chr7	12536838	12542864	+	No		PA3
chr1	178283915	178289953	-	Yes	86	PA2
chr13	94142772	94148790	-	Yes	110	PA2
chr14	60413223	60419241	-	Yes	130	PA2
chr18	8057451	8063461	-	No		PA2
chr2	157111033	157117044	+	Yes	110	PA2

Table S4.1. Endogenous L1 loci expressed in HEK 293 cells contain stop codons in their ORF1 sequence. The top 50 endogenous L1 loci expressed in HEK 293 cells (Deininger et al., 2016 in press) were analyzed for the presence or absence of stop codons within the ORF1 sequence. Columns:

CHR-chromosome #

Left/Right-genomic coordinates based on the hg19 build in the UCSC genome browser

Strand-relative orientation of the L1 locus in the genome, sense (+) or antisense (-)

Stop-presence (Yes) or Absence (No) of a stop codon

Position-position of the first stop codon within the ORF1 protein

Subfamily-indicates L1 subfamily of the L1 locus

NSC-no start codon identified

Table S4.2

CHR	Left	Right	Strand	Stop	Position	Subfamily
					aa	
chr10	27707201	27713319	+	Yes	50	PA4
chr1	121316090	121322219	+	Yes	36	PA3
chr8	48397434	48403568	+	Yes	50	PA5
chr9	72821546	72827671	-	Yes	78	PA5
chr10	102166090	102172241	+	Yes	49	PA4
chr8	70755547	70761658	+	Yes	49	PA5
chr10	95687461	95693583	+	Yes	5	PA5
chr14	55337083	55343223	+	Yes	NSC	PA4
chr6	80287374	80293483	+	Yes	42	PA6
chr7	12536838	12542864	+	No		PA3
chr8	48229550	48235690	+	Yes	73	PA4
chr9	4531072	4537200	+	Yes	49	PA4
chr15	31067384	31073593	+	Yes	NSC	PA4
chr15	90177550	90183532	+	Yes	49	PA3
chr1	53470335	53476454	+	Yes	5	PA5
chr1	149923965	149930070	-	Yes	98	PA5
chr2	32325537	32331652	-	Yes	34	PA5
chr4	20740249	20746401	+	Yes	137	PA5
chr8	54693104	54699226	-	Yes	130	PA4
chr9	108325457	108331580	+	No		PA5
chr15	67740568	67746697	-	Yes	49	PA4
chr1	43580199	43586164	-	Yes	42	PA2
chr2	38359550	38365734	-	Yes	49	PA4
chr3	142193948	142200067	-	Yes	49	PA5
chr8	106801318	106807462	-	No		PA5
chr11	14682876	14688993	+	Yes	50	PA3
chr12	479779	485884	-	Yes	NSC	PA5
chr12	31703001	31709139	-	Yes	12	PA4
chr12	73541398	73547425	-	Yes	130	PA2
chr13	92933026	92939137	+	Yes	31	PA5
chr20	25161484	25167620	-	Yes	NSC	PA6
chr1	51370559	51376681	-	Yes	NSC	PA5
chr1	142626736	142632863	-	Yes	3	PA3
chr1	178806693	178812801	+	Yes	166	PA2
chr1	218182570	218188594	-	No		HS
chr2	73687582	73693778	+	Yes	49	PA4
chr2	180616664	180622826	-	Yes	98	PA5
chr2	202155268	202161393	+	Yes	49	PA4

chr3	15433015	15439121	-	Yes	130	PA4
chr3	17654384	17660525	-	Yes	32	PA6
chr3	38446892	38453029	+	Yes	49	PA3
chr3	51681878	51687951	+	Yes	48	PA6
chr4	49171851	49177978	-	Yes	110	PA3
chr4	57197421	57203548	-	Yes	141	PA3
chr4	67323054	67329193	-	Yes	NSC	PA6
chr4	73380552	73386710	-	Yes	98	PA3
chr4	79897593	79903738	-	Yes	98	PA4
chr4	106781288	106787417	+	Yes	60	PA4
chr5	60171716	60177827	-	Yes	NSC	PA4
chr5	100515835	100521851	+	Yes	130	PA2

Table S4.2. Endogenous L1 loci expressed in HeLa cells contain stop codons in their ORF1 sequence.
The top 50 endogenous L1 loci expressed in HeLa cells (Deininger et al., 2016 in press) were analyzed for the presence or absence of stop codons within the ORF1 sequence. Columns:

CHR-chromosome #

Left/Right-genomic coordinates based on the hg19 build in the UCSC genome browser

Strand-relative orientation of the L1 locus in the genome, sense (+) or antisense (-)

Stop-presence (Yes) or Absence (No) of a stop codon

Position-position of the first stop codon within the ORF1 protein

Subfamily-indicates L1 subfamily of the L1 locus

NSC-no start codon identified

Supplementary Materials and Methods

Plasmids

For subcloning purposes, PCR amplification of all codon optimized DNA sequences, except for C-C, was done using a 5' CCAAGCTTGCCACCATGGGCAAGAAGCAGA 3' forward primer and:

54co Reverse: 5' TTGGATCCTTAGCTGTAGTTGCTGCGGCGGA 3',

98co Reverse: 5' TTGGATCCTTAGGCCTTGGTCTTCAGCTCCATCA 3',

132co Reverse: 5' TTGGATCCTTACTCGCGCTTCATCT 3',

157co Reverse: 5' TTGGATCCTTAGG GGCGCTTCACGTAGTCCCAT 3',

255co Reverse: 5' TTGGATCCTTAGCTCAGGTCGGCGGT 3'.

To generate the C-C construct expressing a T7-tagged truncated ORF1 protein containing only the coiled-coil domain, two rounds of PCR amplification were performed using a C-C Forward 5' CCAAGCTTGCCACCATGAGCGAGCTGCGCGA 3' and a C-C Reverse 5' CCACCAGTCATGCTAGCCA TGGGGCGCTTCACGCAGTCCCAT 3'primer pair for the 1st PCR, then a C-C Forward 5' CCAAGCTTGCCACCATGAGCGAGCTGCGCGA 3' and a C-C Reverse T7 tag 5'GGATCCTTAACCCATTTGCTGTCCACCAGTCATGCTAGCCATGG 3' primer pair for the 2nd PCR.

PCR amplification of all wild-type (non-codon-optimized) DNA sequences was done using a

5' CCAAGCTTGCCGCCACCATGGGGAAAAAACAGAACAGAAAAAC 3' forward primer and:

98wt Reverse: 5' TTGGATCCTTAAGCCTTGGTTTTCAGCTCCATCAG 3',

132wt Reverse: 5' TTGGATCCTTA TTCTCGCTTCATTTCATTCATTTC 3',

157wt Reverse: 5' TTGGATCCTTATGGTCTTTTCACATAGTC CCATA 3'.

To generate the m251 construct expressing a T7-tagged truncated mouse ORF1 protein, two rounds of PCR amplification were performed using m251 Forward: 5'-CCAAGCTTGCCACCATGGCCAAGGGC AAGCGCCGCA-3', m251 Reverse 5'-CCACCAGTCATGCTAGCCAGGCCTCCTGGATGTTCATG TGCATC-3' primer pair for the 1st PCR, then m251 Forward: 5'-CCAAGCTTGCCACCATGGCCAAGGGC AAGCGCCGCA-3' and m251 Reverse T7 tag 5'-GGATCCTTAACCCATTTGCTGTCCACCAGTCATGCTAGCCAT-3' primer pair for the 2nd PCR.

Different human wild-type L1Stop constructs were generated using XL site-directed mutagenesis kit (Agilent) and the following primers:

110Stop: 110Stop F 5' GCAGAAGCCTCAGGAGCTGATGCGATCAACTGGAAG 3', 110Stop R 5' CTTCCAGTTGATCGCATCAGCTCCTGAGGCTTCTGC 3'.

130Stop: 130Stop F 5’ GGAAGATGAAATGAATGAAATGAAGTGAGAAGGGAAGTTTAGAGAAAAAAG 3’, 130Stop R 5’ CTTTTTTCTCTAAACTTCCCTTCTCACTTCATTTCATTCATTTCATCTTCC 3’.

132Stop: 132Stop F 5’ GAAATGAATGAAATGAAGCGAGAATAGAAGTTTAGAGAAAAAAGAATAAAAAG 3’, 132Stop R 5’ CTTTTTATTCTTTTTTCTCTAAACTTCTATTCTCGCTTCATTTCATTCATTTC 3’.

The L1PA1 130 and the L1PA2 130 constructs designed to express truncated L1PA1 and 2 ORF1 proteins were generated with the following primers: forward 5’ CCAAGCTTGCCACCATGATGGGCAAGAAGCAGAACCG 3’, reverse 5’ TTGGATCCTTACTTCATCTCGTTCATCTCGTC 3’.

The L1PA3 130 construct was generated with the following primers: forward 5’ CCAAGCTTGCCACCATGATGGGCAAGAAGCAGAGCCGG 3’, reverse 5’ TTGGATCCTTACTTCATCTCGTTCATCTCGTC 3’.

The ORF1F construct designed to express a full-length ORF1p containing a 3X Flag-tag at its C-terminus was generated by performing 3 rounds of PCR amplification using a forward primer 5’ CCAAGCTTGCCACCATGGGCAAGAAGCAGA 3’ and a reverse primer 5’ ATCACCGTCATGGTCTTTGTAGTCCATCTTGGCGTGGTTTTGCAGGG 3’ for the 1st PCR reaction, then two sequential PCR reactions containing the above mentioned forward primer combined with a reverse primer 5’ ATCGTCATCCTTGTAATCGATGTCATGATCTTTATAATCACCGTCATGGTCTTT 3’, then with a reverse primer 5’ TTGGATCCTTACTTGTCATCGTCATCCTTGTAATC 3’. The resulting PCR product was subcloned into the pBud expression plasmid using HindIII and BamHI restriction endonucleases.

Figures

Figure 1A

L1 retrotransposition assay: $5x10^5$ HeLa cells per T75 flask were seeded 16-18 hours prior to transfection. The cells were cotransfected with 0.4µg of L1Neo plasmid and 0.4µg of either a control (pBud), hORF1 or mORF1 plasmid using 4µl of Plus reagent in the total volume of 200µl of serum free media. After 10 minutes of incubation at room temperature, 8µl of Lipofectamine mixed with 92µl of serum free DMEM/High Glucose media was added to the reaction. The transfection cocktail was incubated for 15 minutes at room temperature and transferred into individual flasks containing 6mL of serum-free DMEM/High Glucose media. 3 hours post transfection, the media was replaced with 10mL of serum containing media. 400µg/mL of G418 (Geneticin, Invitrogen: 10131-027) was administered 24 hours post transfection and maintained for up to 14 days with media changes every 2-3 days. The flasks were stained with 3mL of crystal

violet staining solution (0.2% crystal violet (Sigma C6158), 5% acetic acid (Fisher Scientific A38-212), 2.5% isopropanol (Fisher Scientific BP2632-4)) per flask.

Figure 1B

L1 retrotransposition assay: $7x10^5$ NIH-3T3 cells per T75 flask were seeded 16-18 hours prior to transfection. The cells were cotransfected with 3µg of L1Neo and 0.4µg of either a control (pBud), hORF1 or mORF1 plasmid using 8µl of Plus reagent in the total volume of 200µl of serum free media. After 10 minutes of incubation at room temperature, 12µl of Lipofectamine mixed with 88µl of serum-free DMEM/High Glucose media was added to the reaction. The transfection cocktail was incubated for 15 minutes at room temperature and transferred into individual flasks containing 6mL of serum-free DMEM/High Glucose media. 3 hours post transfection, the media was replaced with 10mL of serum containing media. 300µg/mL of G418 (Geneticin, Invitrogen: 10131-027) was administered 24 hours post transfection and maintained for up to 14 days with media changes every 2-3 days. The flasks were stained with 3mL of crystal violet staining solution (0.2% crystal violet (Sigma C6158), 5% acetic acid (Fisher Scientific A38-212), 2.5% isopropanol (Fisher Scientific BP2632-4)) per flask.

Figure 2B

L1 retrotransposition assay: $5x10^5$ HeLa cells per T75 flask were seeded 16-18 hours prior to transfection. The cells were cotransfected with 0.2µg of L1co Neo plasmid (H1H2, H1M2, M1H2 or M1M2) and 0.4µg of either a control (pBud), hORF1 or mORF1 plasmid using 4µl of Plus reagent in the total volume of 200µl of serum free media. After 10 minutes of incubation at room temperature, 8µl of Lipofectamine mixed with 92µl of serum-free DMEM/High Glucose media was added to the reaction. The transfection cocktail was incubated for 15 minutes at room temperature and transferred into individual flasks containing 6mL of serum-free DMEM/High Glucose media. 3 hours post transfection, the media was replaced with 10mL of serum containing media. 400µg/mL of G418 (Geneticin, Invitrogen: 10131-027) was administered 24 hours post transfection and maintained for up to 14 days with media changes every 2-3 days. The flasks were stained with 3mL of crystal violet staining solution (0.2% crystal violet (Sigma C6158), 5% acetic acid (Fisher Scientific A38-212), 2.5% isopropanol (Fisher Scientific BP2632-4)) per flask.

Figure 2C

L1 retrotransposition assay: $7x10^5$ NIH-3T3 cells per T75 flask were seeded 16-18 hours prior to transfection. The cells were cotransfected with 3.2µg of L1co Neo plasmid (H1H2, H1M2, M1H2 or M1M2) and 3.2µg of either a control (pBud), hORF1 or mORF1 plasmid using 8µl of Plus reagent in the total volume of 200µl of serum free media. After 10 minutes of incubation at room temperature, 12µl of Lipofectamine mixed with 88µl of serum-free DMEM/High Glucose media was added to the reaction. The transfection cocktail was incubated for 15 minutes at room temperature and transferred into individual flasks containing 6mL of serum-free DMEM/High

Glucose media. 3 hours post transfection, the media was replaced with 10mL of serum containing media. 400µg/mL of G418 (Geneticin, Invitrogen: 10131-027) was administered 24 hours post transfection and maintained for up to 14 days with media changes every 2-3 days. The flasks were stained with 3mL of crystal violet staining solution (0.2% crystal violet (Sigma C6158), 5% acetic acid (Fisher Scientific A38-212), 2.5% isopropanol (Fisher Scientific BP2632-4)) per flask.

Figure 3B

Transfection: $7x10^5$ HeLa cells per T25 flask were seeded 16-18 hours prior to transfection. The cells were transfected with 1µg of either a control (pBud), 54co, 98co, 132co, 157co, 255co, hORF1 or C-C plasmid using 4µl of Plus reagent in the total volume of 200µl of serum free media. After 10 minutes of incubation at room temperature, 8µl of Lipofectamine mixed with 92µl of serum free DMEM/High Glucose media was added to the reaction. The transfection cocktail was incubated for 15 minutes at room temperature and transferred into individual flasks containing 2mL of serum-free DMEM/High Glucose media. 3 hours post transfection, the media was replaced with 5mL of serum containing media. Cells were harvested approximately 24 hours later.

Protein harvest for western blot analysis: Cells were washed with 1X PBS (137mM NaCl (Sigma S9888), 2.7nM KCl (Sigma P4505), 10mM Na_2HPO_4 (Sigma S3264) and 2mM KH_2PO_4 (Sigma P9791), pH=7.4) and harvested using 300µl of TLB SDS (50mM Tris, 150mM NaCl, 10mM EDTA, 0.5% sodium dodecyl sulfate, TritonX-100 0.5% v/v, Halt Protease inhibitor 10µl/mL, phosphatase inhibitors 2 and 3 (Sigma), pH=7.2) per T25 flask. The total lysate samples were sonicated three times for 10 seconds at 12 watts RMS using a 3mm wide Qsonica Microson homogenizer with Microson ultrasonic cell disruptor XL2000 (Microson). Samples were centrifuged at 21130xg for 15 minutes at 4-degrees Celsius. The resulting supernatant was transferred to a new microcentrifuge tube. The protein concentration for each sample was determined using 595nm wavelength OD values against a Bovine Serum Albumin (BSA) standard.

Western blot analysis: 35µg of protein for each sample was combined with 2x Laemmli buffer to obtain the final concentration of 1X, 1.6µl β-mercaptoethanol and heated at 100-degrees Celsius for 5 minutes prior to loading. Samples were fractionated on a Bis-Tris 4-12% Midi gel (Invitrogen) and transferred to a nitrocellulose membrane (iBlot2 system: Invitrogen). The membrane was first incubated for 1 hour in the 5% milk/PBS-Tween buffer (0.1% v/v Tween 20 (Sigma P2287), 137mM NaCl (Sigma S9888), 2.7nM KCl (Sigma P4505), 10mM Na_2HPO_4 (Sigma S3264) and 2mM KH_2PO_4 (Sigma P9791), pH=7.4), and then overnight at 4-degrees Celsius with either 1:10000 dilution of T7-tag antibody (Cell Signaling: D9E1X) or with 1:5000 dilution of hORF1 (custom polyclonal rabbit antibody: TGNSKTQSASPPPK (31)) antibody in 3% milk/PBS-Tween buffer. Following the overnight incubation, the membranes were washed for 5 minutes 3 times with PBS-Tween buffer, and incubated at room temperature for 1 hour with 1:5000 dilution of

horseradish peroxidase-conjugated secondary antibodies (HRP-donkey anti-rabbit (Santa Cruz: sc2317)) in 3% milk/PBS-Tween buffer. The membranes were washed one time for 15 minutes with PBS-Tween buffer then twice for 5 minutes with PBS-Tween buffer. The development was done using a 5-minute incubation with a Bio-Rad Clarity Kit (Bio-Rad). The signal was detected using a Chemi-Doc XRS+ Molecular Imager (Bio-Rad). anti-GAPDH antibodies (Santa Cruz: sc25778, 1:5000 dilution) were used to detect GAPDH using incubation conditions described above.

Figure 3C

L1 retrotransposition assay: $5x10^5$ HeLa cells per T75 flask were seeded 16-18 hours prior to transfection. The cells were cotransfected with 0.4µg of L1Neo plasmid and 0.4µg of either a control (pBud), hORF1, mORF1, 54co, 98co, 132co, 157co, 255co or C-C plasmids using 4µl of Plus reagent in the total volume of 200µl of serum-free media. After 10 minutes of incubation at room temperature, 8µl of Lipofectamine mixed with 92µl of serum-free DMEM/High Glucose media was added to the reaction. The transfection cocktail was incubated for 15 minutes at room temperature and transferred into individual flasks containing 6mL of serum-free DMEM/High Glucose media. 3 hours post transfection, the media was replaced with 10mL of serum containing media. 400µg/mL of G418 (Geneticin, Invitrogen: 10131-027) was administered 24 hours post transfection and maintained for up to 14 days with media changes every 2-3 days. The flasks were stained with 3mL of crystal violet staining solution (0.2% crystal violet (Sigma C6158), 5% acetic acid (Fisher Scientific A38-212), 2.5% isopropanol (Fisher Scientific BP2632-4)) per flask.

Figure 3D

Transfection: $7x10^5$ HeLa cells per T25 flask were seeded 16-18 hours prior to transfection. The cells were cotransfected with 1µg of a control (pBud), 98co, 132co, 157co, 255co or hORF1 plasmids and 1µg of either pBud or hORF1 plasmid using 4µl of Plus reagent in the total volume of 200µl of serum-free media. After 10 minutes of incubation at room temperature, 8µl of Lipofectamine mixed with 92µl of serum-free DMEM/High Glucose media was added to the reaction. The transfection cocktail was incubated for 15 minutes at room temperature and transferred into individual flasks containing 2mL of serum-free DMEM/High Glucose media. 3 hours post transfection, the media was replaced with 5mL of serum containing media. Cells were harvested approximately 24 hours later.

Protein harvest for western blot analysis: Cells were washed with 1X PBS (137mM NaCl (Sigma S9888), 2.7nM KCl (Sigma P4505), 10mM Na_2HPO_4 (Sigma S3264) and 2mM KH_2PO_4 (Sigma P9791), pH=7.4) and harvested using 300µl of TLB SDS (50mM Tris, 150mM NaCl, 10mM EDTA, 0.5% sodium dodecyl sulfate, TritonX-100 0.5% v/v, Halt Protease inhibitor 10µl/mL, phosphatase inhibitors 2 and 3 (Sigma), pH=7.2) per T25 flask. The total lysate samples were sonicated five times for 5 seconds at 12 watts RMS using a 3mm wide Qsonica Microson

homogenizer with Microson ultrasonic cell disruptor XL2000 (Microson). Samples were centrifuged at 21130xg for 15 minutes at 4-degrees Celsius. The resulting supernatant was transferred to a new microcentrifuge tube. The protein concentration for each sample was determined using 595nm wavelength OD values against a Bovine Serum Albumin (BSA) standard.

Western blot analysis performed under non-reducing conditions: 25µg of protein for each sample was combined with 2x loading buffer (4% SDS (Sigma L3771), 20% glycerol (v/v) (Sigma G5516), 0.004% bromophenol blue (Bio-Rad 161-0404), 0.125M Tris HCl (Sigma T3253), pH=6.8) to obtain the final concentration of 1X and incubated at 85-degrees Celsius for 5 minutes prior to loading. Samples were fractionated on a Tris-Glycine 4-15% Midi gel (Bio-Rad) and transferred to a nitrocellulose membrane (iBlot2 system: Invitrogen). The membrane was first incubated for 1 hour in the 5% milk/PBS-Tween buffer (0.1% v/v Tween 20 (Sigma P2287), 137mM NaCl (Sigma S9888), 2.7nM KCl (Sigma P4505), 10mM Na_2HPO_4 (Sigma S3264) and 2mM KH_2PO_4 (Sigma P9791), pH=7.4), and then overnight at 4-degrees Celsius with 1:5000 dilution of hORF1 (custom polyclonal rabbit antibody: TGNSKTQSASPPPK) antibody in 3% milk/PBS-Tween buffer. Following the overnight incubation, the membrane was washed 3 times for 5 minutes a piece with PBS-Tween buffer, and incubated at room temperature for 1 hour with 1:5000 dilution of horseradish peroxidase-conjugated secondary antibodies (HRP-donkey anti-rabbit (Santa Cruz: sc2317)) in the 3% milk/PBS-Tween buffer. The membrane was washed one time for 15 minutes with PBS-Tween buffer then twice for 5 minutes with PBS-Tween buffer. The development was done using a 5-minute incubation with a Bio-Rad Clarity Kit (Bio-Rad). The signal was detected using a Chemi-Doc XRS+ Molecular Imager (Bio-Rad). anti-GAPDH antibodies (Santa Cruz: sc25778, 1:5000 dilution) were used to detect GAPDH using incubation conditions described above.

Figure 4B

Transfection: $7x10^5$ HeLa cells per T25 flask were seeded 16-18 hours prior to transfection. The cells were cotransfected with 1µg of either control (pBud), 132co or 132M plasmids and 1µg of either pBud or hORF1 plasmid using 4µl of Plus reagent in the total volume of 200µl of serum-free media. After 10 minutes of incubation at room temperature, 8µl of Lipofectamine mixed with 92µl of serum-free DMEM/High Glucose media was added to the reaction. The transfection cocktail was incubated for 15 minutes at room temperature and transferred into individual flasks containing 2mL of serum-free DMEM/High Glucose media. 3 hours post transfection, the media was replaced with 5mL of serum containing media. Cells were harvested approximately 24 hours later.

Protein harvest for western blot analysis: Cells were washed with 1X PBS (137mM NaCl (Sigma S9888), 2.7nM KCl (Sigma P4505), 10mM Na_2HPO_4 (Sigma S3264) and 2mM KH_2PO_4 (Sigma

P9791), pH=7.4) and harvested using 300µl of TLB lysis buffer (50mM Tris, 150mM NaCl, 10mM EDTA, TritonX-100 0.5% v/v, Halt Protease inhibitor 10µl/mL, phosphatase inhibitors 2 and 3 (Sigma), pH=7.2) per T25 flask. The total lysate samples were sonicated five times for 5 seconds at 12 watts RMS using a 3mm wide Qsonica Microson homogenizer with Microson ultrasonic cell disruptor XL2000 (Microson). Samples were centrifuged at 21130xg for 15 minutes at 4-degrees Celsius. The resulting supernatant was transferred to a new microcentrifuge tube. The protein concentration for each sample was determined using 595nm wavelength OD values against a Bovine Serum Albumin (BSA) standard.

Western blot analysis performed under non-reducing conditions: 22µg of protein for each sample was combined with 2x loading buffer to obtain the final concentration of 1X (4% SDS (Sigma L3771), 20% glycerol (v/v) (Sigma G5516), 0.004% bromophenol blue (Bio-Rad 161-0404), 0.125M Tris HCl (Sigma T3253), pH=6.8) to the final concentration of 1X and incubated at 85-degrees Celsius for 5 minutes prior to loading. Samples were fractionated on a Tris-Glycine 4-15% Midi gel (Bio-Rad) and transferred to a nitrocellulose membrane (iBlot2 system: Invitrogen). The membrane was first incubated for 1 hour in the 5% milk/PBS-Tween buffer (0.1% v/v Tween 20 (Sigma P2287), 137mM NaCl (Sigma S9888), 2.7nM KCl (Sigma P4505), 10mM Na_2HPO_4 (Sigma S3264) and 2mM KH_2PO_4 (Sigma P9791), pH=7.4), and then overnight at 4-degrees Celsius with 1:5000 dilution of hORF1 (custom polyclonal rabbit antibody: TGNSKTQSASPPPK) antibody in the 3% milk/PBS-Tween buffer. Following the overnight incubation, the membrane was washed 3 times for 5 minutes a piece with PBS-Tween buffer, and incubated at room temperature for 1 hour with 1:5000 dilution of horseradish peroxidase-conjugated secondary antibodies (HRP-donkey anti-rabbit (Santa Cruz: sc2317)) in the 3% milk/PBS-Tween buffer. The membrane was washed one time for 15 minutes with PBS-Tween buffer then twice for 5 minutes with PBS-Tween buffer. The development was done using a 5-minute incubation with a Bio-Rad Clarity Kit (Bio-Rad). The signal was detected using a Chemi-Doc XRS+ Molecular Imager (Bio-Rad). anti-GAPDH antibodies (Santa Cruz: sc25778, 1:5000 dilution) were used to detect GAPDH using incubation conditions described above.

Figure 4C

Transfection: $2x10^6$ HeLa cells per T75 flask were seeded 16-18 hours prior to transfection. The cells were cotransfected with 3µg of either a control (pBud), ORF1 Flag, 132co or 132M plasmids and 3µg of either pBud or ORF1 Flag plasmid using 12µl of Plus reagent in the total volume of 200µl of serum-free media. After 10 minutes of incubation at room temperature, 24µl of Lipofectamine mixed with 76µl of serum-free DMEM/High Glucose media was added to the reaction. The transfection cocktail was incubated for 15 minutes at room temperature and transferred into individual flasks containing 6mL of serum-free DMEM/High Glucose media. 3 hours post transfection, the media was replaced with 8mL of serum containing media. Cells were harvested approximately 24 hours post transfection.

Protein harvest for Co-immunoprecipitation: Cells were washed with 1X PBS (137mM NaCl (Sigma S9888), 2.7nM KCl (Sigma P4505), 10mM Na_2HPO_4 (Sigma S3264) and 2mM KH_2PO_4 (Sigma P9791), pH=7.4) and harvested using 500µl of TLB lysis buffer (50mM Tris, 150mM NaCl, 10mM EDTA, TritonX-100 0.5% v/v, Halt Protease inhibitor 10µl/mL, phosphatase inhibitors 2 and 3 (Sigma), pH=7.2) per T75 flask. Samples were centrifuged at 18407xg for 15 minutes at 4-degrees Celsius. The resulting supernatant was transferred to a new microcentrifuge tube (input). The following steps were carried out at 4-degree Celsius. 40µl of flag resin (Anti-Flag M2 Affinity Gel Sigma A2220) was centrifuged at 8200xg for 30 seconds, and 1 minute incubation prior to removing the supernatant of the resin. The resin was washed twice using 500µl of TBS (50mM Tris HCl,150mM NaCl, 1mM EDTA, 1% Triton X-100 v/v, pH=7.4). 200µl of the protein sample was combined with 800µl of TBS buffer incubated with the prepared flag resin overnight at 4-degrees Celsius on a revolving platform. The following day, the mixture was centrifuged at 8200xg for 30 seconds. The supernatant was removed and the resin was washed three times with 500µl of TBS.

After the washes, the remaining protein was eluted by incubation at 100-degrees Celsius for 3 minutes in 20µl of 2X sample buffer (125mM Tris HCl, 4% SDS, 20% (v/v) glycerol, 0.004% bromophenol blue, pH=6.8). The eluted samples were centrifuged for 30 seconds at 8200xg and the supernatant was used for western blot analysis (the co-IP fraction).

Western blot analysis: 10µl of the input for each sample was combined with 10ul of the 2x Laemmli buffer, 1.6µl β-mercaptoethanol and heated at 100-degrees Celsius for 3 minutes prior to loading. The entire volume of the co-IP fraction was used for western blot analysis. Samples were fractionated on a Tris-Acetate 3-8% Midi gel (Invitrogen) and transferred to a nitrocellulose membrane (iBlot2 system: Invitrogen). The membrane was first incubated for 1 hour in the 5% milk/PBS-Tween buffer (0.1% v/v Tween 20 (Sigma P2287), 137mM NaCl (Sigma S9888), 2.7nM KCl (Sigma P4505), 10mM Na_2HPO_4 (Sigma S3264) and 2mM KH_2PO_4 (Sigma P9791), pH=7.4), and then overnight at 4-degrees Celsius with 1:5000 dilution of hORF1 (custom polyclonal rabbit antibody: TGNSKTQSASPPPK) antibody in the 3% milk/PBS-Tween buffer. Following the overnight incubation, the membrane was washed 3 times for 5 minutes a piece with PBS-Tween buffer, and incubated at room temperature for 1 hour with 1:5000 dilution of horseradish peroxidase-conjugated secondary antibodies (HRP-donkey anti-rabbit (Santa Cruz: sc2317)) in 3% milk in PBS-Tween buffer. The membrane was washed one time for 15 minutes with PBS-Tween buffer then twice for 5 minutes with PBS-Tween buffer. The development was done using a 5-minute incubation with a Bio-Rad Clarity Kit (Bio-Rad). The signal was detected using a Chemi-Doc XRS+ Molecular Imager (Bio-Rad). anti-GAPDH antibodies (Santa Cruz: sc25778, 1:5000 dilution) were used to detect GAPDH using incubation conditions described above.

Figure 4D

Transfection: 2x10^6 HeLa cells per T75 flask were seeded 16-18 hours prior to transfection. The cells were cotransfected with 1µg of either control (pBud), hORF1, 132co or 132M plasmids and either 1µg of pBud or hORF1 plasmids using 6µl of Plus reagent in the total volume of 200µl of serum-free media. After 10 minutes of incubation at room temperature, 12µl of Lipofectamine mixed with 88µl of serum-free DMEM/High Glucose media was added to the reaction. The transfection cocktail was incubated for 15 minutes at room temperature and transferred into individual flasks containing 6mL of serum-free DMEM/High Glucose media. 3 hours post transfection, the media was replaced with 8mL of serum containing media. Cells were harvested approximately 24 hours later.

Protein harvest for western blot analysis: Cells were washed with 1X PBS (137mM NaCl (Sigma S9888), 2.7nM KCl (Sigma P4505), 10mM Na_2HPO_4 (Sigma S3264) and 2mM KH_2PO_4 (Sigma P9791), pH=7.4) and harvested using 500µl of TLB lysis buffer (50mM Tris, 150mM NaCl, 10mM EDTA, TritonX-100 0.5% v/v, Halt Protease inhibitor 10µl/mL, phosphatase inhibitors 2 and 3 (Sigma), pH=7.2) per T75 flask. Samples were centrifuged at 21130xg for 15 minutes at 4-degrees Celsius. The resulting supernatant was used for western blot analysis (cytoplasmic fraction). The resulting pellet was resuspended using 200µl of TLB SDS (50mM Tris, 150mM NaCl, 10mM EDTA, 0.5% sodium dodecyl sulfate, TritonX-100 0.5% v/v, Halt Protease inhibitor 10µl/mL, phosphatase inhibitors 2 and 3 (Sigma), pH=7.2) and sonicated three times for 10 seconds at 12 watts RMS using a 3mm wide Qsonica Microson homogenizer with Microson ultrasonic cell disruptor XL2000 (Microson). The samples were centrifuged at 21130xg for 15 minutes at 4-degrees Celsius. The resulting supernatant was used for western blot analysis (nuclear fraction). The protein concentration for each sample was determined using 595nm wavelength OD values against a Bovine Serum Albumin (BSA) standard.

Western blot analysis: 25µg of protein for each sample was combined with 2x Laemmli buffer to final concentration of 1X, 1.6µl β-mercaptoethanol and heated at 100-degrees Celsius for 5 minutes prior to loading. Samples were fractionated on a Bis-Tris 4-12% Midi gel (Invitrogen) and transferred to a nitrocellulose membrane (iBlot2 system: Invitrogen). The membrane was first incubated at room temperature for 1 hour in the 5% milk/PBS-Tween buffer (0.1% v/v Tween 20 (Sigma P2287), 137mM NaCl (Sigma S9888), 2.7nM KCl (Sigma P4505), 10mM Na_2HPO_4 (Sigma S3264) and 2mM KH_2PO_4 (Sigma P9791), pH=7.4), then overnight at 4-degrees Celsius with 1:5000 dilution of hORF1 (custom polyclonal rabbit antibody: TGNSKTQSASPPPK) antibody in the 3% milk/PBS-Tween buffer. Following the overnight incubation, the membrane was washed 3 times for 5 minutes a piece with PBS-Tween buffer, and incubated at room temperature for 1 hour with 1:5000 dilution of horseradish peroxidase-conjugated secondary antibodies (HRP-donkey anti-rabbit (Santa Cruz: sc2317)) in the 3% milk/PBS-Tween buffer. The membrane was washed one time for 15 minutes with PBS-Tween buffer then twice for 5 minutes with PBS-Tween buffer. The development was done using a 5-minute incubation with a Bio-Rad Clarity Kit (Bio-

Rad). The signal was detected using a Chemi-Doc XRS+ Molecular Imager (Bio-Rad). anti-GAPDH antibodies (Santa Cruz: sc25778, 1:5000 dilution) were used to detect GAPDH using incubation conditions described above. anti-Lamin A/C antibodies (Santa Cruz 7293, 1:1000 dilution) were used to detect Lamin A/C using incubation conditions described above with HRP-goat anti-mouse (Santa Cruz; sc-2031)) at 1:5,000 dilution.

Figure 4E

L1 retrotransposition assay: $5x10^5$ HeLa cells per T75 flask were seeded 16-18 hours prior to transfection. The cells were cotransfected with 0.4µg of L1Neo plasmid and 0.4µg of either a control (pBud), hORF1, mORF1, 132co or 132M plasmid using 4µl of Plus reagent in the total volume of 200µl of serum free media. After 10 minutes of incubation at room temperature, 8µl of Lipofectamine mixed with 92µl of serum-free DMEM/High Glucose media was added to the reaction. The transfection cocktail was incubated for 15 minutes at room temperature and transferred into individual flasks containing 6mL of serum-free DMEM/High Glucose media. 3 hours post transfection, the media was replaced with 10mL of serum containing media. 400µg/mL of G418 (Geneticin, Invitrogen: 10131-027) was administered 24 hours post transfection and maintained for up to 14 days with media changes every 2-3 days. The flasks were stained with 3mL of crystal violet staining solution (0.2% crystal violet (Sigma C6158), 5% acetic acid (Fisher Scientific A38-212), 2.5% isopropanol (Fisher Scientific BP2632-4)) per flask.

Figure 5B

Transfection: $7x10^5$ HeLa cells per T25 flask were seeded 16-18 hours prior to transfection. The cells were transfected with 1µg of either a control (pBud), 98co, 132co, 157co, 98wt, 132wt or 157wt plasmid using 4µl of Plus reagent in the total volume of 200µl of serum-free media. After 10 minutes of incubation at room temperature, 8µl of Lipofectamine mixed with 92µl of serum-free DMEM/High Glucose media was added to the reaction. The transfection cocktail was incubated for 15 minutes at room temperature and transferred into individual flasks containing 2mL of serum-free DMEM/High Glucose media. 3 hours post transfection, the media was replaced with 5mL of serum containing media. Cells were harvested approximately 24 hours later.

Protein harvest for western blot analysis: Cells were washed with 1X PBS (137mM NaCl (Sigma S9888), 2.7nM KCl (Sigma P4505), 10mM Na_2HPO_4 (Sigma S3264) and 2mM KH_2PO_4 (Sigma P9791), pH=7.4) and harvested using 300µl of TLB SDS (50mM Tris, 150mM NaCl, 10mM EDTA, 0.5% sodium dodecyl sulfate, TritonX-100 0.5% v/v, Halt Protease inhibitor 10µl/mL, phosphatase inhibitors 2 and 3 (Sigma), pH=7.2) per T25 flask. The total lysate samples were sonicated three times for 10 seconds at 12 watts RMS using a 3mm wide Qsonica Microson homogenizer with Microson ultrasonic cell disruptor XL2000 (Microson). Samples were centrifuged at 21130xg for 15 minutes at 4-degrees Celsius. The resulting supernatant was transferred to a new microcentrifuge tube. The protein concentration for each sample was

determined using 595nm wavelength OD values against a Bovine Serum Albumin (BSA) standard.

Western blot analysis: 28µg of protein for each sample was combined with 2x Laemmli buffer to obtain the final concentration of 1X, 1.6µl β-mercaptoethanol and heated at 100-degrees Celsius for 5 minutes prior to loading. Samples were fractionated on a Bis-Tris 4-12% Midi gel (Invitrogen) and transferred to a nitrocellulose membrane (iBlot2 system: Invitrogen). The membrane was first incubated for 1 hour in the 5% milk/PBS-Tween buffer (0.1% v/v Tween 20 (Sigma P2287), 137mM NaCl (Sigma S9888), 2.7nM KCl (Sigma P4505), 10mM Na_2HPO_4 (Sigma S3264) and 2mM KH_2PO_4 (Sigma P9791), pH=7.4), and then overnight at 4-degrees Celsius with 1:5000 dilution of hORF1 (custom polyclonal rabbit antibody: TGNSKTQSASPPPK) antibody in the 3% milk/PBS-Tween buffer. Following the overnight incubation, the membrane was washed 3 times for 5 minutes a piece with PBS-Tween buffer, and incubated at room temperature for 1 hour with 1:5000 dilution of horseradish peroxidase-conjugated secondary antibodies (HRP-donkey anti-rabbit (Santa Cruz: sc2317)) in 3% milk in PBS-Tween buffer. The membrane was washed one time for 15 minutes with PBS-Tween buffer then twice for 5 minutes with PBS-Tween buffer. The development was done using a 5-minute incubation with a Bio-Rad Clarity Kit (Bio-Rad). The signal was detected using a Chemi-Doc XRS+ Molecular Imager (Bio-Rad). anti-GAPDH antibodies (Santa Cruz: sc25778, 1:5000 dilution) were used to detect GAPDH using incubation conditions described above.

Figure 5C

L1 retrotransposition assay: $5x10^5$ HeLa cells per T75 flask were seeded 16-18 hours prior to transfection. The cells were cotransfected with 0.4µg of L1Neo plasmid and 0.4µg of either a control (pBud), hORF1, mORF1, 98wt, 132wt or 157wt plasmid using 4µl of Plus reagent in the total volume of 200µl of serum-free media. After 10 minutes of incubation at room temperature, 8µl of Lipofectamine mixed with 92µl of serum-free DMEM/High Glucose media was added to the reaction. The transfection cocktail was incubated for 15 minutes at room temperature and transferred into individual flasks containing 6mL of serum-free DMEM/High Glucose media. 3 hours post transfection, the media was replaced with 10mL of serum containing media. 400µg/mL of G418 (Geneticin, Invitrogen: 10131-027) was administered 24 hours post transfection and maintained for up to 14 days with media changes every 2-3 days. The flasks were stained with 3mL of crystal violet staining solution (0.2% crystal violet (Sigma C6158), 5% acetic acid (Fisher Scientific A38-212), 2.5% isopropanol (Fisher Scientific BP2632-4)) per flask.

Figure 6A

L1Base (51) was used to identify full-length L1 loci containing stop codons in their ORF1 sequence. Human Full-Length >4,500nt LINE-1 Elements Ens38.36 data set was used to identify the L1 loci containing no gaps or frameshifts in ORF1 and one or more stop codons in ORF1 with

a requirement for the query start sites within 50 base pairs relative to the start of the 5'UTR of a full-length L1 element. This search identified 1474 L1 loci, which were then analyzed using the amino acid FASTA output for each of the ORF1 sequences to confirm the presence of a methionine residue corresponding to the beginning of the human ORF1p. This approach identified 1244 L1 loci containing ORF1p with a start codon. These L1 loci were further analyzed to identify the position of the first stop codon using their amino acid FASTA sequence. The frequency of stop codons at each ORF1 amino acid position was plotted (Figure 6A).

Figure 6B

Transfection: $7x10^5$ HeLa cells per T25 flask were seeded 16-18 hours prior to transfection. The cells were transfected with 1µg of either a control (pBud), L1PA1 FL, L1PA2 FL, L1PA3 FL, L1PA1 130, L1PA2 130, L1PA3 130, e207, e127 or e259 plasmid using 4µl of Plus reagent in the total volume of 200µl of serum-free media. After 10 minutes of incubation at room temperature, 8µl of Lipofectamine mixed with 92µl of serum-free DMEM/High Glucose media was added to the reaction. The transfection cocktail was incubated for 15 minutes at room temperature and transferred into individual flasks containing 2mL of serum-free DMEM/High Glucose media. 3 hours post transfection, the media was replaced with 5mL of serum containing media. Cells were harvested approximately 24 hours later.

Protein harvest for western blot analysis: Cells were washed with 1X PBS (137mM NaCl (Sigma S9888), 2.7nM KCl (Sigma P4505), 10mM Na_2HPO_4 (Sigma S3264) and 2mM KH_2PO_4 (Sigma P9791), pH=7.4) and harvested using 300µl of TLB SDS (50mM Tris, 150mM NaCl, 10mM EDTA, 0.5% sodium dodecyl sulfate, TritonX-100 0.5% v/v, Halt Protease inhibitor 10µl/mL, phosphatase inhibitors 2 and 3 (Sigma), pH=7.2) per T25 flask. The total lysate samples were sonicated three times for 10 seconds at 12 watts RMS using a 3mm wide Qsonica Microson homogenizer with Microson ultrasonic cell disruptor XL2000 (Microson). Samples were centrifuged at 21130xg for 15 minutes at 4-degrees Celsius. The resulting supernatant was transferred to a new microcentrifuge tube. The protein concentration for each sample was determined using 595nm wavelength OD values against a Bovine Serum Albumin (BSA) standard.

Western blot analysis: 40µg of protein for each sample was combined with 2x Laemmli buffer to obtain the final concentration of 1X, 1.6µl β-mercaptoethanol and heated at 100-degrees Celsius for 5 minutes prior to loading. Samples were fractionated on a Bis-Tris 4-12% Midi gel (Invitrogen) and transferred to a nitrocellulose membrane (iBlot2 system: Invitrogen). The membrane was first incubated for 1 hour in the 5% milk/PBS-Tween buffer (0.1% v/v Tween 20 (Sigma P2287), 137mM NaCl (Sigma S9888), 2.7nM KCl (Sigma P4505), 10mM Na_2HPO_4 (Sigma S3264) and 2mM KH_2PO_4 (Sigma P9791), pH=7.4), and then overnight at 4-degrees Celsius with 1:5000 dilution of hORF1 (custom polyclonal rabbit antibody: TGNSKTQSASPPPK) antibody in the 3%

milk/PBS-Tween buffer. Following the overnight incubation, the membrane was washed 3 times for 5 minutes a piece with PBS-Tween buffer, and incubated at room temperature for 1 hour with 1:5000 dilution of horseradish peroxidase-conjugated secondary antibodies (HRP-donkey anti-rabbit (Santa Cruz: sc2317)) in 3% milk in PBS-Tween buffer. The membrane was washed one time for 15 minutes with PBS-Tween buffer then twice for 5 minutes with PBS-Tween buffer. The development was done using a 5-minute incubation with a Bio-Rad Clarity Kit (Bio-Rad). The signal was detected using a Chemi-Doc XRS+ Molecular Imager (Bio-Rad). anti-GAPDH antibodies (Santa Cruz: sc25778, 1:5000 dilution) were used to detect GAPDH using incubation conditions described above.

Figure 6C

L1 retrotransposition assay: $5x10^5$ HeLa cells per T75 flask were seeded 16-18 hours prior to transfection. The cells were cotransfected with 0.4µg of L1Neo plasmid and 0.4µg of either a control (pBud), L1PA1 FL, L1PA2 FL, L1PA3 FL, L1PA1 130, L1PA2 130, L1PA3 130, e207, e127 or e259 plasmid using 4µl of Plus reagent in the total volume of 200µl of serum-free media. After 10 minutes of incubation at room temperature, 8µl of Lipofectamine mixed with 92µl of serum-free DMEM/High Glucose media was added to the reaction. The transfection cocktail was incubated for 15 minutes at room temperature and transferred into individual flasks containing 6mL of serum-free DMEM/High Glucose media. 3 hours post transfection, the media was replaced with 10mL of serum containing media. 400µg/mL of G418 (Geneticin, Invitrogen: 10131-027) was administered 24 hours post transfection and maintained for up to 14 days with media changes every 2-3 days. The flasks were stained with 3mL of crystal violet staining solution (0.2% crystal violet (Sigma C6158), 5% acetic acid (Fisher Scientific A38-212), 2.5% isopropanol (Fisher Scientific BP2632-4)) per flask.

Figure 7B

Transfection: $7x10^5$ HeLa cells per T25 flask were seeded 16-18 hours prior to transfection. The cells were transfected with 4µg of either a control (pCEP), hL1wt, 110Stop, 119Stop, 130Stop or 132Stop plasmid using 6µl of Plus reagent in the total volume of 200µl of serum-free media. After 10 minutes of incubation at room temperature, 12µl of Lipofectamine mixed with 88µl of serum-free DMEM/High Glucose media was added to the reaction. The transfection cocktail was incubated for 15 minutes at room temperature and transferred into individual flasks containing 2mL of serum-free DMEM/High Glucose media. 3 hours post transfection, the media was replaced with 5mL of serum containing media. Cells were harvested approximately 24 hours later.

Protein harvest for western blot analysis: Cells were washed with 1X PBS (137mM NaCl (Sigma S9888), 2.7nM KCl (Sigma P4505), 10mM Na_2HPO_4 (Sigma S3264) and 2mM KH_2PO_4 (Sigma P9791), pH=7.4) and harvested using 300µl of TLB SDS (50mM Tris, 150mM NaCl, 10mM EDTA, 0.5% sodium dodecyl sulfate, TritonX-100 0.5% v/v, Halt Protease inhibitor 10µl/mL,

phosphatase inhibitors 2 and 3 (Sigma), pH=7.2) per T25 flask. The total lysate samples were sonicated three times for 10 seconds at 12 watts RMS using a 3mm wide Qsonica Microson homogenizer with Microson ultrasonic cell disruptor XL2000 (Microson). Samples were centrifuged at 21130xg for 15 minutes at 4-degrees Celsius. The resulting supernatant was transferred to a new microcentrifuge tube. The protein concentration for each sample was determined using 595nm wavelength OD values against a Bovine Serum Albumin (BSA) standard.

Western blot analysis: 12µg of protein for each sample was combined with 2x Laemmli buffer to obtain the final concentration of 1X, 1.6µl β-mercaptoethanol and heated at 100-degrees Celsius for 5 minutes prior to loading. Samples were fractionated on a Bis-Tris 4-12% Midi gel (Invitrogen) and transferred to a nitrocellulose membrane (iBlot2 system: Invitrogen). The membrane was first incubated for 1 hour in the 5% milk/PBS-Tween buffer (0.1% v/v Tween 20 (Sigma P2287), 137mM NaCl (Sigma S9888), 2.7nM KCl (Sigma P4505), 10mM Na_2HPO_4 (Sigma S3264) and 2mM KH_2PO_4 (Sigma P9791), pH=7.4), and then overnight at 4-degrees Celsius with 1:5000 dilution of hORF1 (custom polyclonal rabbit antibody: TGNSKTQSASPPPK) antibody in the 3% milk/PBS-Tween buffer. Following the overnight incubation, the membrane was washed 3 times for 5 minutes a piece with PBS-Tween buffer, and incubated at room temperature for 1 hour with 1:5000 dilution of horseradish peroxidase-conjugated secondary antibodies (HRP-donkey anti-rabbit (Santa Cruz: sc2317)) in 3% milk in PBS-Tween buffer. The membrane was washed one time for 15 minutes with PBS-Tween buffer then twice for 5 minutes with PBS-Tween buffer. The development was done using a 5-minute incubation with a Bio-Rad Clarity Kit (Bio-Rad). The signal was detected using a Chemi-Doc XRS+ Molecular Imager (Bio-Rad). anti-GAPDH antibodies (Santa Cruz: sc25778, 1:5000 dilution) were used to detect GAPDH using incubation conditions described above.

Figure 7C

L1 retrotransposition assay: $5x10^5$ HeLa cells per T75 flask were seeded 16-18 hours prior to transfection. The cells were cotransfected with 0.4µg of L1Neo plasmid and 0.4µg of either a control (pCEP), hL1wt, 110Stop, 119Stop, 130Stop, 132Stop or 5'UTR plasmid using 4µl of Plus reagent in the total volume of 200µl of serum-free media. After 10 minutes of incubation at room temperature, 8µl of Lipofectamine mixed with 92µl of serum-free DMEM/High Glucose media was added to the reaction. The transfection cocktail was incubated for 15 minutes at room temperature and transferred into individual flasks containing 6mL of serum-free DMEM/High Glucose media. 3 hours post transfection, the media was replaced with 10mL of serum containing media. 400µg/mL of G418 (Geneticin, Invitrogen: 10131-027) was administered 24 hours post transfection and maintained for up to 14 days with media changes every 2-3 days. The flasks were stained with 3mL of crystal violet staining solution (0.2% crystal violet (Sigma C6158), 5% acetic acid (Fisher Scientific A38-212), 2.5% isopropanol (Fisher Scientific BP2632-4)) per flask.

Figure 8B

Generation of engineered HeLa cell lines harboring plasmids expressing L1wt or L1Stop elements: $5x10^5$ HeLa cells per T75 flask were seeded 16-18 hours prior to transfection. The cells were transfected with 1µg of either a control (pCEP), hL1wt, 119Stop, 130Stop or 132Stop plasmids using 4µl of Plus reagent in the total volume of 200µl of serum-free media. After 10 minutes of incubation at room temperature, 8µl of Lipofectamine mixed with 92µl of serum-free DMEM/High Glucose media was added to the reaction. The transfection cocktail was incubated for 15 minutes at room temperature and transferred into individual flasks containing 6mL of serum-free DMEM/High Glucose media. 3 hours post transfection, the media was replaced with 10mL of serum containing media. Hygromycin B (Invitrogen: 10687-010) selection was administered 24 hours post transfection with final concentration of Hygromycin B in the selection media of 225µg/mL. Each constitutively expressing HeLa flask was fed and passaged under Hygromycin B selection for 14 days before seeding for experiments.

Protein harvest for western blot analysis: $2x10^6$ cells of each engineered HeLa lines were seeded per a T75 flask. Total protein was harvested 40 hours after seeding. Cells were washed with 1X PBS (137mM NaCl (Sigma S9888), 2.7nM KCl (Sigma P4505), 10mM Na_2HPO_4 (Sigma S3264) and 2mM KH_2PO_4 (Sigma P9791), pH=7.4) and harvested using 500µl of TLB SDS (50mM Tris, 150mM NaCl, 10mM EDTA, 0.5% sodium dodecyl sulfate, TritonX-100 0.5% v/v, Halt Protease inhibitor 10µl/mL, phosphatase inhibitors 2 and 3 (Sigma), pH=7.2) per T75 flask. The total lysate samples were sonicated three times for 10 seconds at 12 watts RMS using a 3mm wide Qsonica Microson homogenizer with Microson ultrasonic cell disruptor XL2000 (Microson). Samples were centrifuged at 21130xg for 15 minutes at 4-degrees Celsius. The resulting supernatant was transferred to a new microcentrifuge tube. The protein concentration for each sample was determined using 595nm wavelength OD values against a Bovine Serum Albumin (BSA) standard.

Western blot analysis: 22µg of protein for each sample was combined with 2x Laemmli buffer to obtain the final concentration of 1X, 1.6µl β-mercaptoethanol and heated at 100-degrees Celsius for 5 minutes prior to loading. Samples were fractionated on a Bis-Tris 4-12% Midi gel (Invitrogen) and transferred to a nitrocellulose membrane (iBlot2 system: Invitrogen). The membrane was first incubated for 1 hour in the 5% milk/PBS-Tween buffer (0.1% v/v Tween 20 (Sigma P2287), 137mM NaCl (Sigma S9888), 2.7nM KCl (Sigma P4505), 10mM Na_2HPO_4 (Sigma S3264) and 2mM KH_2PO_4 (Sigma P9791), pH=7.4), and then overnight at 4-degrees Celsius with 1:5000 dilution of hORF1 (custom polyclonal rabbit antibody: TGNSKTQSASPPPK) antibody in the 3% milk/PBS-Tween buffer. Following the overnight incubation, the membrane was washed 3 times for 5 minutes a piece with PBS-Tween buffer, and incubated at room temperature for 1 hour with 1:5000 dilution of horseradish peroxidase-conjugated secondary antibodies (HRP-donkey anti-rabbit (Santa Cruz: sc2317)) in 3% milk in PBS-Tween buffer. The membrane was washed one

time for 15 minutes with PBS-Tween buffer then twice for 5 minutes with PBS-Tween buffer. The development was done using a 5-minute incubation with a Bio-Rad Clarity Kit (Bio-Rad). The signal was detected using a Chemi-Doc XRS+ Molecular Imager (Bio-Rad). anti-GAPDH antibodies (Santa Cruz: sc25778, 1:5000 dilution) were used to detect GAPDH using incubation conditions described above.

Figure 8C

L1 retrotransposition assay: $5x10^5$ of engineered HeLa cells per T75 flask were seeded 16-18 hours prior to transfection. The cells were transfected with 0.1µg of L1Neo plasmid using 4µl of Plus reagent in the total volume of 200µl of serum-free media. After 10 minutes of incubation at room temperature, 8µl of Lipofectamine mixed with 92µl of serum-free DMEM/High Glucose media was added to the reaction. The transfection cocktail was incubated for 15 minutes at room temperature and transferred into individual flasks containing 6mL of serum-free DMEM/High Glucose media. 3 hours post transfection, the media was replaced with 10mL of serum containing media. 400µg/mL of G418 (Geneticin, Invitrogen: 10131-027) was administered 24 hours post transfection and maintained for up to 14 days with media changes every 2-3 days. The flasks were stained with 3mL of crystal violet staining solution (0.2% crystal violet (Sigma C6158), 5% acetic acid (Fisher Scientific A38-212), 2.5% isopropanol (Fisher Scientific BP2632-4)) per flask.

Transfection efficiency control: $5x10^5$ engineered HeLa cells per T75 flask were seeded 16-18 hours prior to transfection. The cells were transfected with 0.1µg of pIRES2-GFP plasmid using 4µl of Plus reagent in the total volume of 200µl of serum-free media. After 10 minutes of incubation at room temperature, 8µl of Lipofectamine mixed with 92µl of serum-free DMEM/High Glucose media was added to the reaction. The transfection cocktail was incubated for 15 minutes at room temperature and transferred into individual flasks containing 6mL of serum-free DMEM/High Glucose media. 3 hours post transfection, the media was replaced with 10mL of serum containing media. 400µg/mL of G418 (Geneticin, Invitrogen: 10131-027) was administered 24 hours post transfection and maintained for up to 14 days with media changes every 2-3 days. The flasks were stained with 3mL of crystal violet staining solution (0.2% crystal violet (Sigma C6158), 5% acetic acid (Fisher Scientific A38-212), 2.5% isopropanol (Fisher Scientific BP2632-4)) per flask**.**

Figure S4.2B

Toxicity assay: $5x10^5$ HeLa cells per T75 flask were seeded 16-18 hours prior to transfection. The cells were cotransfected with 0.1µg of pIRES2-GFP and 0.4µg of either a control (pBud), hORF1, mORF1, 132co or 132M plasmid using 4µl of Plus reagent in the total volume of 200µl of serum-free media. After 10 minutes of incubation at room temperature, 8µl of Lipofectamine mixed with 92µl of serum-free DMEM/High Glucose media was added to the reaction. The transfection cocktail was incubated for 15 minutes at room temperature and transferred into individual flasks

containing 6mL of serum-free DMEM/High Glucose media. 3 hours post transfection, the media was replaced with 10mL of serum containing media. 400µg/mL of G418 (Geneticin, Invitrogen: 10131-027) was administered 24 hours post transfection and maintained for up to 14 days with media changes every 2-3 days. The flasks were stained with 3mL of crystal violet staining solution (0.2% crystal violet (Sigma C6158), 5% acetic acid (Fisher Scientific A38-212), 2.5% isopropanol (Fisher Scientific BP2632-4)) per flask.

Figure S4.2C

Toxicity assay: $5x10^5$ HeLa cells per T75 flask were seeded 16-18 hours prior to transfection. The cells were cotransfected with 0.1µg of pIRES2-GFP and 0.4µg of hL1wt plasmid and 0.4µg of either a control (pBud), hORF1, mORF1, 132co or 132M plasmid using 4µl of Plus reagent in the total volume of 200µl of serum-free media. After 10 minutes of incubation at room temperature, 8µl of Lipofectamine mixed with 92µl of serum-free DMEM/High Glucose media was added to the reaction. The transfection cocktail was incubated for 15 minutes at room temperature and transferred into individual flasks containing 6mL of serum-free DMEM/High Glucose media. 3 hours post transfection, the media was replaced with 10mL of serum containing media. 400µg/mL of G418 (Geneticin, Invitrogen: 10131-027) was administered 24 hours post transfection and maintained for up to 14 days with media changes every 2-3 days. The flasks were stained with 3mL of crystal violet staining solution (0.2% crystal violet (Sigma C6158), 5% acetic acid (Fisher Scientific A38-212), 2.5% isopropanol (Fisher Scientific BP2632-4)) per flask.

Figure S4.3

Transfection: $7x10^5$ HeLa cells per T25 flask were seeded 16-18 hours prior to transfection. The cells were cotransfected with 1µg of control plasmid (pBud), C-C or hORF1 plasmid, and either 1µg of pBud or hORF1 plasmid using 4µl of Plus reagent in the total volume of 200µl of serum-free media. After 10 minutes of incubation at room temperature, 8µl of Lipofectamine mixed with 92µl of serum-free DMEM/High Glucose media was added to the reaction. The transfection cocktail was incubated for 15 minutes at room temperature and transferred into individual flasks containing 2mL of serum-free DMEM/High Glucose media. 3 hours post transfection, the media was replaced with 5mL of serum containing media. Cells were harvested approximately 24 hours later.

Protein harvest for western blot analysis: Cells were washed with 1X PBS (137mM NaCl (Sigma S9888), 2.7nM KCl (Sigma P4505), 10mM Na_2HPO_4 (Sigma S3264) and 2mM KH_2PO_4 (Sigma P9791), pH=7.4) and harvested using 300µl of TLB lysis buffer (50mM Tris, 150mM NaCl, 10mM EDTA, TritonX-100 0.5% v/v, Halt Protease inhibitor 10µl/mL, phosphatase inhibitors 2 and 3 (Sigma), pH=7.2) per T25 flask. The total lysate samples were sonicated five times for 5 seconds at 12 watts RMS using a 3mm wide Qsonica Microson homogenizer with Microson ultrasonic cell disruptor XL2000 (Microson). Samples were centrifuged at 21130xg for 15 minutes at 4-degrees

Celsius. The resulting supernatant was transferred to a new microcentrifuge tube. The protein concentration for each sample was determined using 595nm wavelength OD values against a Bovine Serum Albumin (BSA) standard.

Western blot analysis performed under non-reducing conditions: 25µg of protein for each sample was combined with 2x loading buffer to obtain the final concentration of 1X (4% SDS (Sigma L3771), 20% glycerol (v/v) (Sigma G5516), 0.004% bromophenol blue (Bio-Rad 161-0404), 0.125M Tris HCl (Sigma T3253), pH=6.8) and incubated at 85-degrees Celsius for 5 minutes prior to loading. Samples were fractionated on a Tris-Glycine 4-15% Midi gel (Bio-Rad) and transferred to a nitrocellulose membrane (iBlot2 system: Invitrogen). The membrane was first incubated for 1 hour in the 5% milk/PBS-Tween buffer (0.1% v/v Tween 20 (Sigma P2287), 137mM NaCl (Sigma S9888), 2.7nM KCl (Sigma P4505), 10mM Na_2HPO_4 (Sigma S3264) and 2mM KH_2PO_4 (Sigma P9791), pH=7.4), and then overnight at 4-degrees Celsius with 1:5000 dilution of hORF1 (custom polyclonal rabbit antibody: TGNSKTQSASPPPK) antibody in the 3% milk/PBS-Tween buffer. Following the overnight incubation, the membrane was washed 3 times for 5 minutes a piece with PBS-Tween buffer, and incubated at room temperature for 1 hour with 1:5000 dilution of horseradish peroxidase-conjugated secondary antibodies (HRP-donkey anti-rabbit (Santa Cruz: sc2317)) in 3% milk in PBS-Tween buffer. The membrane was washed one time for 15 minutes with PBS-Tween buffer then twice for 5 minutes with PBS-Tween buffer. The development was done using a 5-minute incubation with a Bio-Rad Clarity Kit (Bio-Rad). The signal was detected using a Chemi-Doc XRS+ Molecular Imager (Bio-Rad). anti-GAPDH antibodies (Santa Cruz: sc25778, 1:5000 dilution) were used to detect GAPDH using incubation conditions described above.

Figure S4.4A

Transfection: $7x10^5$ HeLa cells per T25 flask were seeded 16-18 hours prior to transfection. The cells were cotransfected with 1µg of control plasmid (pBud), hNmC-C or mNhC-C plasmid, and either 1µg of pBud or hORF1 plasmid using 4µl of Plus reagent in the total volume of 200µl of serum-free media. After 10 minutes of incubation at room temperature, 8µl of Lipofectamine mixed with 92µl of serum-free DMEM/High Glucose media was added to the reaction. The transfection cocktail was incubated for 15 minutes at room temperature and transferred into individual flasks containing 2mL of serum-free DMEM/High Glucose media. 3 hours post transfection, the media was replaced with 5mL of serum containing media. Cells were harvested approximately 24 hours later.

Protein harvest for western blot analysis: Cells were washed with 1X PBS (137mM NaCl (Sigma S9888), 2.7nM KCl (Sigma P4505), 10mM Na_2HPO_4 (Sigma S3264) and 2mM KH_2PO_4 (Sigma P9791), pH=7.4) and harvested using 300µl of TLB SDS (50mM Tris, 150mM NaCl, 10mM EDTA, 0.5% sodium dodecyl sulfate, TritonX-100 0.5% v/v, Halt Protease inhibitor 10µl/mL,

phosphatase inhibitors 2 and 3 (Sigma), pH=7.2) per T25 flask. The total lysate samples were sonicated three times for 10 seconds at 12 watts RMS using a 3mm wide Qsonica Microson homogenizer with Microson ultrasonic cell disruptor XL2000 (Microson). Samples were centrifuged at 21130xg for 15 minutes at 4-degrees Celsius. The resulting supernatant was transferred to a new microcentrifuge tube. The protein concentration for each sample was determined using 595nm wavelength OD values against a Bovine Serum Albumin (BSA) standard.

Western blot analysis: 30µg of protein for each sample was combined with 2x Laemmli buffer to obtain the final concentration of 1X, 1.6µl β-mercaptoethanol and heated at 100-degrees Celsius for 5 minutes prior to loading. Samples were fractionated on a Bis-Tris 4-12% Midi gel (Invitrogen) and transferred to a nitrocellulose membrane (iBlot2 system: Invitrogen). The membrane was first incubated for 1 hour in the 5% milk/PBS-Tween buffer (0.1% v/v Tween 20 (Sigma P2287), 137mM NaCl (Sigma S9888), 2.7nM KCl (Sigma P4505), 10mM Na_2HPO_4 (Sigma S3264) and 2mM KH_2PO_4 (Sigma P9791), pH=7.4), and then overnight at 4-degrees Celsius with T7-tag antibody (Cell Signaling; D9E1X, 1:10000 dilution) in 3% milk in PBS-Tween buffer. Following the overnight incubation, the membrane was washed 3 times for 5 minutes a piece with PBS-Tween buffer, and incubated at room temperature for 1 hour with 1:5000 dilution of horseradish peroxidase-conjugated secondary antibodies (HRP-donkey anti-rabbit (Santa Cruz: sc2317)) in 3% milk in PBS-Tween buffer. The membrane was washed one time for 15 minutes with PBS-Tween buffer then twice for 5 minutes with PBS-Tween buffer. The development was done using a 5-minute incubation with a Bio-Rad Clarity Kit (Bio-Rad). The signal was detected using a Chemi-Doc XRS+ Molecular Imager (Bio-Rad). anti-GAPDH antibodies (Santa Cruz: sc25778, 1:5000 dilution) were used to detect GAPDH using incubation conditions described above.

Figure S4.4B

Transfection: $7x10^5$ HeLa cells per T25 flask were seeded 16-18 hours prior to transfection. The cells were cotransfected with 1µg of either a control (pBud), mNhC-C, or hORF1 plasmid, and then 1µg of either a control (pBud), or hORF1 using 4µl of Plus reagent in the total volume of 200µl of serum-free media. After 10 minutes of incubation at room temperature, 8µl of Lipofectamine mixed with 92µl of serum-free DMEM/High Glucose media was added to the reaction. The transfection cocktail was incubated for 15 minutes at room temperature and transferred into individual flasks containing 2mL of serum-free DMEM/High Glucose media. 3 hours post transfection, the media was replaced with 5mL of serum containing media. Cells were harvested approximately 24 hours later.

Protein harvest for western blot analysis: Cells were washed with 1X PBS (137mM NaCl (Sigma S9888), 2.7nM KCl (Sigma P4505), 10mM Na_2HPO_4 (Sigma S3264) and 2mM KH_2PO_4 (Sigma P9791), pH=7.4) and harvested using 300µl of TLB SDS (50mM Tris, 150mM NaCl, 10mM

EDTA, 0.5% sodium dodecyl sulfate, TritonX-100 0.5% v/v, Halt Protease inhibitor 10µl/mL, phosphatase inhibitors 2 and 3 (Sigma), pH=7.2) per T25 flask. The total lysate samples were sonicated five times for 5 seconds at 12 watts RMS using a 3mm wide Qsonica Microson homogenizer with Microson ultrasonic cell disruptor XL2000 (Microson). Samples were centrifuged at 21130xg for 15 minutes at 4-degrees Celsius. The resulting supernatant was transferred to a new microcentrifuge tube. The protein concentration for each sample was determined using 595nm wavelength OD values against a Bovine Serum Albumin (BSA) standard.

Western blot analysis: 25µg of protein for each sample was combined with 2x Laemmli buffer to obtain the final concentration of 1X, 1.6µl β-mercaptoethanol and heated at 100-degrees Celsius for 5 minutes prior to loading. Samples were fractionated on a Tris-Glycine 4-15% (Bio-Rad) and transferred to a nitrocellulose membrane (iBlot2 system: Invitrogen). The membrane was first incubated for 1 hour in the 5% milk/PBS-Tween buffer (0.1% v/v Tween 20 (Sigma P2287), 137mM NaCl (Sigma S9888), 2.7nM KCl (Sigma P4505), 10mM Na_2HPO_4 (Sigma S3264) and 2mM KH_2PO_4 (Sigma P9791), pH=7.4), and then overnight at 4-degrees Celsius with 1:5000 dilution of hORF1 (custom polyclonal rabbit antibody: TGNSKTQSASPPPK) antibody in the 3% milk/PBS-Tween buffer. Following the overnight incubation, the membrane was washed 3 times for 5 minutes a piece with PBS-Tween buffer, and incubated at room temperature for 1 hour with 1:5000 dilution of horseradish peroxidase-conjugated secondary antibodies (HRP-donkey anti-rabbit (Santa Cruz: sc2317)) in 3% milk in PBS-Tween buffer. The membrane was washed one time for 15 minutes with PBS-Tween buffer then twice for 5 minutes with PBS-Tween buffer. The development was done using a 5-minute incubation with a Bio-Rad Clarity Kit (Bio-Rad). The signal was detected using a Chemi-Doc XRS+ Molecular Imager (Bio-Rad). anti-GAPDH antibodies (Santa Cruz: sc25778, 1:5000 dilution) were used to detect GAPDH using incubation conditions described above.

Figure S4.4C

L1 retrotransposition assay: $5x10^5$ HeLa cells per T75 flask were seeded 16-18 hours prior to transfection. The cells were cotransfected with 0.4µg of L1Neo plasmid and 0.4µg of either a control (pBud), hORF1, mORF1, mNhC-C or hNmC-C plasmid using 4µl of Plus reagent in the total volume of 200µl of serum-free media. After 10 minutes of incubation at room temperature, 8µl of Lipofectamine mixed with 92µl of serum-free DMEM/High Glucose media was added to the reaction. The transfection cocktail was incubated for 15 minutes at room temperature and transferred into individual flasks containing 6mL of serum-free DMEM/High Glucose media. 3 hours post transfection, the media was replaced with 10mL of serum containing media. 400µg/mL of G418 (Geneticin, Invitrogen: 10131-027) was administered 24 hours post transfection and maintained for up to 14 days with media changes every 2-3 days. The flasks were stained with

3mL of crystal violet staining solution (0.2% crystal violet (Sigma C6158), 5% acetic acid (Fisher Scientific A38-212), 2.5% isopropanol (Fisher Scientific BP2632-4)) per flask.

Figure S4.5

Transfection: $7x10^5$ HeLa cells per T25 flask were seeded 16-18 hours prior to transfection. The cells were cotransfected with 1µg of control plasmid (pBud), ORF1, ORF1 Flag or Gal4 tagged ORF1 plasmid, and 1µg of either pBud or hORF1 Flag plasmid using 4µl of Plus reagent in the total volume of 200µl of serum-free media. After 10 minutes of incubation at room temperature, 8µl of Lipofectamine mixed with 92µl of serum-free DMEM/High Glucose media was added to the reaction. The transfection cocktail was incubated for 15 minutes at room temperature and transferred into individual flasks containing 2mL of serum-free DMEM/High Glucose media. 3 hours post transfection, the media was replaced with 5mL of serum containing media. Cells were harvested approximately 24 hours later.

Protein harvest for western blot analysis: Cells were washed with 1X PBS. Cells were harvested using 300µl of TLB SDS (50mM Tris, 150mM NaCl, 10mM EDTA, 0.5% sodium dodecyl sulfate, TritonX-100 0.5% v/v, Halt Protease inhibitor 10µl/mL, phosphatase inhibitors 2 and 3 (Sigma), pH=7.2) per T25 flask. The total lysate samples were sonicated five times for 5 seconds at 12 watts RMS using a 3mm wide Qsonica Microson homogenizer with Microson ultrasonic cell disruptor XL2000 (Microson). Samples were centrifuged at 21130xg for 15 minutes at 4-degrees Celsius. The supernatant of the samples was transferred to a new microcentrifuge tube. The protein concentrations of the samples were determined using 595nm wavelength OD values against a Bovine Serum Albumin (BSA) standard.

Western blot analysis performed under non-reducing conditions: 20µg of protein for each sample was combined with 2x loading buffer to obtain the final concentration of 1X (4% SDS (Sigma L3771), 20% glycerol (v/v) (Sigma G5516), 0.004% bromophenol blue (Bio-Rad 161-0404), 0.125M Tris HCl (Sigma T3253), pH=6.8) and incubated at 85-degrees Celsius for 5 minutes prior to loading. Samples were fractionated on a Tris-Acetate 3-8% Midi gel (Invitrogen) and transferred to a nitrocellulose membrane (iBlot2 system: Invitrogen). The membrane was first incubated for 1 hour in the 5% milk/PBS-Tween buffer (0.1% v/v Tween 20 (Sigma P2287), 137mM NaCl (Sigma S9888), 2.7nM KCl (Sigma P4505), 10mM Na_2HPO_4 (Sigma S3264) and 2mM KH_2PO_4 (Sigma P9791), pH=7.4), and then overnight at 4-degrees Celsius with 1:5000 dilution of hORF1 (custom polyclonal rabbit antibody: TGNSKTQSASPPPK) antibody in the 3% milk/PBS-Tween buffer. Following the overnight incubation, the membrane was washed for 3 times for 5 minutes a piece with PBS-Tween buffer, and incubated at room temperature for 1 hour with 1:5000 dilution of horseradish peroxidase-conjugated secondary antibodies (HRP-donkey anti-rabbit (Santa Cruz: sc2317)) in 3% milk in PBS-Tween buffer. The membrane was washed one time for 15 minutes with PBS-Tween buffer then twice for 5 minutes with PBS-Tween buffer.

The development was done using a 5-minute incubation with a Bio-Rad Clarity Kit (Bio-Rad). The signal was detected using a Chemi-Doc XRS+ Molecular Imager (Bio-Rad). In parallel analysis, the same samples were run under the same conditions except using a 1:10000 dilution of Flag tag (Sigma Monoclonal Anti-Flag M2: F3165) antibody and then HRP-goat anti-mouse (Santa Cruz; sc-2031)) at 1:5000 dilution. antibody anti-GAPDH antibodies (Santa Cruz: sc25778, 1:5000 dilution) were used to detect GAPDH using incubation conditions described above.

Figure S4.6

See Figure 5B for details. This western blot is the same one shown in Figure 5B but exposed for a longer time to show the detection of truncated ORF1 protein generated from the 157wt and 98wt expression plasmids.

Figure S4.7B

L1 retrotransposition assay 0.2 conditions: $5x10^5$ HeLa cells per T75 flask were seeded 16-18 hours prior to transfection. The cells were cotransfected with 0.4µg of L1Neo plasmid, 0.2µg of either a control (pBud), hORF1, mORF1, 98wt, 157wt, 98co or 157co plasmid, and 0.2µg of pBud, 157wt or 157co plasmid using 4µl of Plus reagent in the total volume of 200µl of serum-free media. After 10 minutes of incubation at room temperature, 8µl of Lipofectamine mixed with 92µl of serum-free DMEM/High Glucose media was added to the reaction. The transfection cocktail was incubated for 15 minutes at room temperature and transferred into individual flasks containing 6mL of serum-free DMEM/High Glucose media. 3 hours post transfection, the media was replaced with 10mL of serum containing media. 400µg/mL of G418 (Geneticin, Invitrogen: 10131-027) was administered 24 hours post transfection and maintained for up to 14 days with media changes every 2-3 days. The flasks were stained with 3mL of crystal violet staining solution (0.2% crystal violet (Sigma C6158), 5% acetic acid (Fisher Scientific A38-212), 2.5% isopropanol (Fisher Scientific BP2632-4)) per flask.

L1 retrotransposition assay 0.4 conditions: $5x10^5$ HeLa cells per T75 flask were seeded 16-18 hours prior to transfection. The cells were cotransfected with 0.4µg of L1Neo plasmid, 0.4µg of either a control (pBud), hORF1, mORF1, 98wt, 157wt, 98co or 157co plasmid and 0.4µg of pBud, 157wt or 157co plasmid using 4µl of Plus reagent in the total volume of 200µl of serum-free media. After 10 minutes of incubation at room temperature, 8µl of Lipofectamine mixed with 92µl of serum-free DMEM/High Glucose media was added to the reaction. The transfection cocktail was incubated for 15 minutes at room temperature and transferred into individual flasks containing 6mL of serum-free DMEM/High Glucose media. 3 hours post transfection, the media was replaced with 10mL of serum containing media. 400µg/mL of G418 (Geneticin, Invitrogen: 10131-027) was administered 24 hours post transfection and maintained for up to 14 days with media changes every 2-3 days. The flasks were stained with 3mL of crystal violet staining

solution (0.2% crystal violet (Sigma C6158), 5% acetic acid (Fisher Scientific A38-212), 2.5% isopropanol (Fisher Scientific BP2632-4)) per flask.

Figure S4.8

L1.3 ORF1 sequence was analyzed for base pair substitutions away from a one of the stop codon positions for each codon. The results are listed in the Supplementary Figure S8.

Figure S4.9

Manual analysis of SNPs in the ORF1 sequence of 50 full-length fixed L1 loci (reference build hg19 of the UCSC genome browser) identified 4 L1 loci in which the presence of stop codons in ORF1 between amino acids 110-130 was polymorphic. The results of the analysis are shown in Supplementary Figure S9.

Figure S4.10

L1Base (51) was utilized to identify full-length L1 loci containing stop codons in their ORF1 sequence. L1Base Mouse Full-Length >5knt LINE-1 Elements Ens38.35 data set was used to identify the L1 loci containing no gaps or frameshifts in ORF1 and one or more stop codons in ORF1 with a requirement for the query start sites within 50 base pairs relative to the start of the 5'UTR of a full-length L1 element. The mouse L1 loci were analyzed using the amino acid FASTA output for each of the ORF1 sequences to confirm the presence of a methionine residue corresponding to the beginning of the mouse ORF1p. The first 10 L1 locus entries identified on each mouse chromosome (a total of 198 loci) were subjected to the first stop codon position analysis. The frequency of stop positions at each amino acid of the mouse ORF1 protein was plotted (Supplementary Figure S10).

Figure S4.11B

Transfection: $5x10^5$ HeLa cells per T75 flask were seeded 16-18 hours prior to transfection. The cells were transfected with 3µg of a control plasmid (pBud) or m251 plasmid using 6µl of Plus reagent in the total volume of 200µl of serum-free media. After 10 minutes of incubation at room temperature, 24µl of Lipofectamine mixed with 76µl of serum-free DMEM/High Glucose media was added to the reaction. The transfection cocktail was incubated for 15 minutes at room temperature and transferred into individual flasks containing 2mL of serum-free DMEM/High Glucose media. 3 hours post transfection, the media was replaced with 5mL of serum containing media. Cells were harvested approximately 24 hours later.

Protein harvest for western blot analysis: Cells were washed with 1X PBS (137mM NaCl (Sigma S9888), 2.7nM KCl (Sigma P4505), 10mM Na_2HPO_4 (Sigma S3264) and 2mM KH_2PO_4 (Sigma P9791), pH=7.4) and harvested using 300µl of TLB SDS (50mM Tris, 150mM NaCl, 10mM EDTA, 0.5% sodium dodecyl sulfate, TritonX-100 0.5% v/v, Halt Protease inhibitor 10µl/mL, phosphatase inhibitors 2 and 3 (Sigma), pH=7.2) per T25 flask. The total lysate samples were

sonicated three times for 10 seconds at 12 watts RMS using a 3mm wide Qsonica Microson homogenizer with Microson ultrasonic cell disruptor XL2000 (Microson). Samples were centrifuged at 21130xg for 15 minutes at 4-degrees Celsius. The resulting supernatant was transferred to a new microcentrifuge tube. The protein concentration for each sample was determined using 595nm wavelength OD values against a Bovine Serum Albumin (BSA) standard.

Western blot analysis: 14µg of protein for each sample was combined with 2x Laemmli buffer to obtain the final concentration of 1X, 1.6µl β-mercaptoethanol and heated at 100-degrees Celsius for 5 minutes prior to loading. Samples were fractionated on a Bis-Tris 4-12% Midi gel (Invitrogen) and transferred to a nitrocellulose membrane (iBlot2 system: Invitrogen). The membrane was first incubated for 1 hour in the 5% milk/PBS-Tween buffer (0.1% v/v Tween 20 (Sigma P2287), 137mM NaCl (Sigma S9888), 2.7nM KCl (Sigma P4505), 10mM Na_2HPO_4 (Sigma S3264) and 2mM KH_2PO_4 (Sigma P9791), pH=7.4) at room temperature, and then overnight at 4-degrees Celsius with 1:5000 dilution of hORF1 (custom polyclonal rabbit antibody: TGNSKTQSASPPPK) antibody in the 3% milk/PBS-Tween buffer. Following the overnight incubation, the membrane was washed 3 times for 5 minutes a piece with PBS-Tween buffer, and incubated for 1 hour at room temperature with 1:5000 dilution of horseradish peroxidase-conjugated secondary antibodies (HRP-donkey anti-rabbit (Santa Cruz: sc2317)) in 3% milk in PBS-Tween buffer. The membrane was washed one time for 15 minutes with PBS-Tween buffer then twice for 5 minutes with PBS-Tween buffer. The development was done using a 5-minute incubation with a Bio-Rad Clarity Kit (Bio-Rad). The signal was detected using a Chemi-Doc XRS+ Molecular Imager (Bio-Rad). In parallel analysis, the same samples were run under the same conditions except using a T7-tag antibody (Cell Signaling; D9E1X, 1:10000 dilution) and then 1:5000 dilution of horseradish peroxidase-conjugated secondary antibodies (HRP-donkey anti-rabbit (Santa Cruz: sc2317). anti-GAPDH antibodies (Santa Cruz: sc25778, 1:5000 dilution) were used to detect GAPDH using incubation conditions described above.

Figure S4.11C

Transfection: $7x10^5$ HeLa cells per T25 flask were seeded 16-18 hours prior to transfection. The cells were cotransfected with 1µg of a control plasmid (pBud), mORF1 or m251 plasmid and 1µg of either pBud or mORF1 plasmid using 4µl of Plus reagent in the total volume of 200µl of serum-free media. After 10 minutes of incubation at room temperature, 8µl of Lipofectamine mixed with 92µl of serum-free DMEM/High Glucose media was added to the reaction. The transfection cocktail was incubated for 15 minutes at room temperature and transferred into individual flasks containing 2mL of serum-free DMEM/High Glucose media. 3 hours post transfection, the media was replaced with 5mL of serum containing media. Cells were harvested approximately 24 hours later.

Protein harvest for western blot analysis: Cells were washed with 1X PBS (137mM NaCl (Sigma S9888), 2.7nM KCl (Sigma P4505), 10mM Na_2HPO_4 (Sigma S3264) and 2mM KH_2PO_4 (Sigma P9791), pH=7.4) and harvested using 300µl of TLB lysis buffer (50mM Tris, 150mM NaCl, 10mM EDTA, TritonX-100 0.5% v/v, Halt Protease inhibitor 10µl/mL, phosphatase inhibitors 2 and 3 (Sigma), pH=7.2) per T25 flask. The total lysate samples were sonicated five times for 5 seconds at 12 watts RMS using a 3mm wide Qsonica Microson homogenizer with Microson ultrasonic cell disruptor XL2000 (Microson). Samples were centrifuged at 21130xg for 15 minutes at 4-degrees Celsius. The resulting supernatant was transferred to a new microcentrifuge tube. The protein concentration for each sample was determined using 595nm wavelength OD values against a Bovine Serum Albumin (BSA) standard.

Western blot analysis performed under non-reducing conditions: 31µg of protein for each sample was combined with 2x loading buffer to obtain the final concentration of 1X (4% SDS (Sigma L3771), 20% glycerol (v/v) (Sigma G5516), 0.004% bromophenol blue (Bio-Rad 161-0404), 0.125M Tris HCl (Sigma T3253), pH=6.8) and incubated at 85-degrees Celsius for 5 minutes prior to loading. Samples were fractionated on a Tris-Glycine 4-15% Midi gel (Bio-Rad) and transferred to a nitrocellulose membrane (iBlot2 system: Invitrogen). The membrane was first incubated for 1 hour in the 5% milk/PBS-Tween buffer (0.1% v/v Tween 20 (Sigma P2287), 137mM NaCl (Sigma S9888), 2.7nM KCl (Sigma P4505), 10mM Na_2HPO_4 (Sigma S3264) and 2mM KH_2PO_4 (Sigma P9791), pH=7.4), and then overnight at 4-degrees Celsius with 1:1000 dilution of anti-mORF1p (epitope: YRTPNRLDQKRNSS) antibody in the 3% milk/PBS-Tween buffer. Following the overnight incubation, the membrane was washed 3 times for 5 minutes a piece with PBS-Tween buffer, and incubated for 1 hour at room temperature with 1:5000 dilution of horseradish peroxidase-conjugated secondary antibodies (HRP-donkey anti-goat (Santa Cruz; sc-2020)) in 3% milk in PBS-Tween buffer. The membrane was washed one time for 15 minutes with PBS-Tween buffer then twice for 5 minutes with PBS-Tween buffer. The development was done using a 5-minute incubation with a Bio-Rad Clarity Kit (Bio-Rad). The signal was detected using a Chemi-Doc XRS+ Molecular Imager (Bio-Rad). In parallel analysis, the same samples were run under the same conditions except using a T7-tag antibody (Cell Signaling; D9E1X, 1:10000 dilution) and then 1:5000 dilution of horseradish peroxidase-conjugated secondary antibodies (HRP-donkey anti-rabbit (Santa Cruz: sc2317). anti-GAPDH antibodies (Santa Cruz: sc25778, 1:5000 dilution) were used to detect GAPDH using incubation conditions described above.

Figure S4.11D

L1 retrotransposition assay: $5x10^5$ HeLa cells per T75 flask were seeded 16-18 hours prior to transfection. The cells were cotransfected with 1µg of mL1wt Neo plasmid and 1µg of either a control (pBud), hORF1, mORF1 or m251 plasmid using 4µl of Plus reagent in the total volume of 200µl of serum-free media. After 10 minutes of incubation at room temperature, 8µl of Lipofectamine mixed with 92µl of serum-free DMEM/High Glucose media was added to the

reaction. The transfection cocktail was incubated for 15 minutes at room temperature and transferred into individual flasks containing 6mL of serum-free DMEM/High Glucose media. 3 hours post transfection, the media was replaced with 10mL of serum containing media. 400µg/mL of G418 (Geneticin, Invitrogen: 10131-027) was administered 24 hours post transfection and maintained for up to 14 days with media changes every 2-3 days. The flasks were stained with 3mL of crystal violet staining solution (0.2% crystal violet (Sigma C6158), 5% acetic acid (Fisher Scientific A38-212), 2.5% isopropanol (Fisher Scientific BP2632-4)) per flask.

Figure S4.12

L1 retrotransposition assay: $5x10^5$ HeLa cells per T75 flask were seeded 16-18 hours prior to transfection. The cells were cotransfected with 0.4µg of L1Neo plasmid and 0.4µg of either a control (pBud), hORF1, mORF1 or m251 plasmid using 4µl of Plus reagent in the total volume of 200µl of serum-free media. After 10 minutes of incubation at room temperature, 8µl of Lipofectamine mixed with 92µl of serum-free DMEM/High Glucose media was added to the reaction. The transfection cocktail was incubated for 15 minutes at room temperature and transferred into individual flasks containing 6mL of serum-free DMEM/High Glucose media. 3 hours post transfection, the media was replaced with 10mL of serum containing media. 400µg/mL of G418 (Geneticin, Invitrogen: 10131-027) was administered 24 hours post transfection and maintained for up to 14 days with media changes every 2-3 days. The flasks were stained with 3mL of crystal violet staining solution (0.2% crystal violet (Sigma C6158), 5% acetic acid (Fisher Scientific A38-212), 2.5% isopropanol (Fisher Scientific BP2632-4)) per flask.

Figure S4.13A and Table 4.1

The top 50 L1 loci identified by PacBio sequencing (Deininger et. al., 2016) (51) to be expressed in HEK 293 cells were analyzed for the presence of a stop codon within their ORF1 sequence using the amino acid FASTA sequence. The results are shown in the Supplementary Table 2. The frequency of identified stop codons at each ORF1 amino acid positions was plotted (Supplementary Figure 13B). The L1 loci subfamily composition is listed in Supplementary Figure 13B.

Figure S4.13B and Table 4.2

The top 50 L1 loci identified by PacBio sequencing (Deininger et. al., 2016) (51) to be expressed in HeLa cells were analyzed for the presence of a stop codon within their ORF1 sequence using the amino acid FASTA sequence. The results are shown in the Supplementary Table 2. The frequency of identified stop codons at each ORF1 amino acid positions was plotted (Supplementary Figure 13B). The L1 loci subfamily composition is listed in Supplementary Figure 13B.

Figure S4.14

Sucrose cushion of the cytoplasmic extract: Briefly, two T75 flasks per cell line were seeded at $4x10^6$ cells 24 hours before harvest. Each T75 flasks was washed 3 times with ice cold 1X phosphate buffered saline (Gibco, 10010023). 5mL of ice cold PBS was added to each flask, the cells were scraped and the two flasks were combined in one 15mL conical tube. The tubes were centrifuged for 5 minutes at 3,000xg at 4 degrees Celsius. The supernatant was removed and 500 microliters of lysis buffer (1.5mM KCl, 2.5mM $MgCl_2$, 5mM Tris-Cl pH=7.5, 1% deoxycholic acid,1% v/v Triton X-100 (Sigma), Halt protease inhibitor 10µl/mL, RNAsin (Promega) 2µl/mL) was added to the pellet, mixed, and incubated on ice for 5 minutes. The sample was centrifuged for 5 minutes at 3,000xg at 4 degrees Celsius. The resulting supernatant was then layered on the top of a sucrose (Sigma) cushion (17% on the bottom (1.5mL) and 8.5% placed carefully on top (1.5mL) (8.5% or 17% sucrose (Sigma), 80mM NaCl, 5mM $MgCl_2$, 20mM Tris-Cl pH=7.5, 1mM DTT, Halt protease inhibitor 10µl/mL and 5µl/mL of RNAsin). The samples were centrifuged for 2 hours at 36500xg at 4 degrees Celsius, after which the pellet was resuspended in 200µl of TLB SDS lysis buffer (50mM Tris, 150mM NaCl, 10mM EDTA, 0.5% sodium dodecyl sulfate, TritonX-100 0.5% v/v, Halt Protease inhibitor 10µl/mL, phosphatase inhibitors 2 and 3 (Sigma), pH=7.2).

The samples were sonicated three times for 10 seconds at 12 watts RMS using a 3mm wide Qsonica Microson homogenizer with Microson ultrasonic cell disruptor XL2000 (Microson). Samples were centrifuged at 21130xg for 15 minutes at 4-degrees Celsius and the resulting supernatant was used for analysis. The protein concentrations of the samples were determined using 595nm wavelength OD values against a Bovine Serum Albumin (BSA) standard.

Western blot analysis: 15µg of each sample was combined with 2x Laemmli buffer to obtain the final concentration of 1X, 1.6µl β-mercaptoethanol and heated at 100-degrees Celsius for 5 minutes prior to loading. The 132co (1.5µg), 255co (3.5µg) and hORF1 (1.5µg) samples (same samples as used in Figure 3B) were used as molecular markers for full-length and truncated ORF1 protein. Samples were fractionated on a Bis-Tris 4-12% Midi gel (Invitrogen) and transferred to a nitrocellulose membrane (iBlot2 system: Invitrogen). The membrane was first incubated for 1 hour in the 5% milk/PBS-Tween buffer (0.1% v/v Tween 20 (Sigma P2287), 137mM NaCl (Sigma S9888), 2.7nM KCl (Sigma P4505), 10mM Na_2HPO_4 (Sigma S3264) and 2mM KH_2PO_4 (Sigma P9791), pH=7.4), and then overnight at 4-degrees Celsius with 1:5000 dilution of hORF1 (custom polyclonal rabbit antibody: TGNSKTQSASPPPK) antibody in the 3% milk/PBS-Tween buffer. Following the overnight incubation, the membrane was washed 3 times for 5 minutes a piece with PBS-Tween buffer, and incubated for 1 hour at room temperature with 1:5000 dilution of horseradish peroxidase-conjugated secondary antibodies (HRP-donkey anti-rabbit (Santa Cruz: sc2317)) in 3% milk in PBS-Tween buffer. The membrane was washed one time for 15 minutes with PBS-Tween buffer then twice for 5 minutes with PBS-Tween buffer. The development was done using a 5-minute incubation with a Bio-Rad Clarity Kit (Bio-Rad). The signal was detected using a Chemi-Doc XRS+ Molecular Imager (Bio-Rad). anti-GAPDH

antibodies (Santa Cruz: sc25778, 1:5000 dilution) were used to detect GAPDH using incubation conditions described above.

Figure S4.15A

Transfection: $7x10^5$ HeLa cells per T25 flask were seeded 16-18 hours prior to transfection. The cells were cotransfected with 1µg of a control plasmid (pBud), Gal4-tagged ORF1, full-length ORF1 PA2 or full-length ORF1 PA3 plasmid and 1µg of either pBud or Gal4-tagged ORF1 plasmid using 4µl of Plus reagent in the total volume of 200µl of serum-free media. After 10 minutes of incubation at room temperature, 8µl of Lipofectamine mixed with 92µl of serum-free DMEM/High Glucose media was added to the reaction. The transfection cocktail was incubated for 15 minutes at room temperature and transferred into individual flasks containing 2mL of serum-free DMEM/High Glucose media. 3 hours post transfection, the media was replaced with 5mL of serum containing media. Cells were harvested approximately 24 hours later.

Protein harvest for western blot analysis: Cells were washed with 1X PBS (137mM NaCl (Sigma S9888), 2.7nM KCl (Sigma P4505), 10mM Na_2HPO_4 (Sigma S3264) and 2mM KH_2PO_4 (Sigma P9791), pH=7.4) and harvested using 300µl of TLB lysis buffer (50mM Tris, 150mM NaCl, 10mM EDTA, TritonX-100 0.5% v/v, Halt Protease inhibitor 10µl/mL, phosphatase inhibitors 2 and 3 (Sigma), pH=7.2) per T25 flask. The total lysate samples were sonicated five times for 5 seconds at 12 watts RMS using a 3mm wide Qsonica Microson homogenizer with Microson ultrasonic cell disruptor XL2000 (Microson). Samples were centrifuged at 21130xg for 15 minutes at 4-degrees Celsius. The resulting supernatant was transferred to a new microcentrifuge tube. The protein concentration for each sample was determined using 595nm wavelength OD values against a Bovine Serum Albumin (BSA) standard.

Western blot analysis performed under non-reducing conditions: 20µg of each sample was combined with 2x loading buffer to obtain the final concentration of 1X (4% SDS (Sigma L3771), 20% glycerol (v/v) (Sigma G5516), 0.004% bromophenol blue (Bio-Rad 161-0404), 0.125M Tris HCl (Sigma T3253), pH=6.8) and incubated at 85-degrees Celsius for 5 minutes prior to loading. Samples were fractionated on a Tris Acetate 3-8% Midi gel (Invitrogen) and transferred to a nitrocellulose membrane (iBlot2 system: Invitrogen). The membrane was first incubated for 1 hour in the 5% milk/PBS-Tween buffer (0.1% v/v Tween 20 (Sigma P2287), 137mM NaCl (Sigma S9888), 2.7nM KCl (Sigma P4505), 10mM Na_2HPO_4 (Sigma S3264) and 2mM KH_2PO_4 (Sigma P9791), pH=7.4), and then overnight at 4-degrees Celsius with 1:5000 dilution of hORF1 (custom polyclonal rabbit antibody: TGNSKTQSASPPPK) antibody in the 3% milk/PBS-Tween buffer. Following the overnight incubation, the membrane was washed 3 times for 5 minutes a piece with PBS-Tween buffer, and incubated for 1 hour at room temperature with 1:5000 dilution of horseradish peroxidase-conjugated secondary antibodies (HRP-donkey anti-rabbit (Santa Cruz: sc2317)) in 3% milk in PBS-Tween buffer. The membrane was washed one time for 15 minutes

with PBS-Tween buffer then twice for 5 minutes with PBS-Tween buffer. The development was done using a 5-minute incubation with a Bio-Rad Clarity Kit (Bio-Rad). The signal was detected using a Chemi-Doc XRS+ Molecular Imager (Bio-Rad). anti-GAPDH antibodies (Santa Cruz: sc25778, 1:5000 dilution) were used to detect GAPDH using incubation conditions described above.

Figure S4.15B

Transfection: $7x10^5$ HeLa cells per T25 flask were seeded 16-18 hours prior to transfection. The cells were cotransfected with 1µg of a control plasmid (pBud), ORF1, ORF1 PA1 130, ORF1 PA2 130 or ORF1 PA3 130 plasmid and 1µg of either pBud or ORF1 plasmid using 4µl of Plus reagent in the total volume of 200µl of serum-free media. After 10 minutes of incubation at room temperature, 8µl of Lipofectamine mixed with 92µl of serum-free DMEM/High Glucose media was added to the reaction. The transfection cocktail was incubated for 15 minutes at room temperature and transferred into individual flasks containing 2mL of serum-free DMEM/High Glucose media. 3 hours post transfection, the media was replaced with 5mL of serum containing media. Cells were harvested approximately 24 hours later.

Protein harvest for western blot analysis: Cells were washed with 1X PBS (137mM NaCl (Sigma S9888), 2.7nM KCl (Sigma P4505), 10mM Na_2HPO_4 (Sigma S3264) and 2mM KH_2PO_4 (Sigma P9791), pH=7.4) and harvested using 300µl of TLB lysis buffer (50mM Tris, 150mM NaCl, 10mM EDTA, TritonX-100 0.5% v/v, Halt Protease inhibitor 10µl/mL, phosphatase inhibitors 2 and 3 (Sigma), pH=7.2) per T25 flask. The total lysate samples were sonicated five times for 5 seconds at 12 watts RMS using a 3mm wide Qsonica Microson homogenizer with Microson ultrasonic cell disruptor XL2000 (Microson). Samples were centrifuged at 21130xg for 15 minutes at 4-degrees Celsius. The resulting supernatant was transferred to a new microcentrifuge tube. The protein concentration for each sample was determined using 595nm wavelength OD values against a Bovine Serum Albumin (BSA) standard.

Western blot analysis performed under non-reducing conditions: 20µg of each sample was combined with 2x loading buffer to obtain the final concentration of 1X (4% SDS (Sigma L3771), 20% glycerol (v/v) (Sigma G5516), 0.004% bromophenol blue (Bio-Rad 161-0404), 0.125M Tris HCl (Sigma T3253), pH=6.8) and incubated at 85-degrees Celsius for 5 minutes prior to loading. Samples were fractionated on a Tris Acetate 3-8% Midi gel (Invitrogen) and transferred to a nitrocellulose membrane (iBlot2 system: Invitrogen). The membrane was first incubated for 1 hour in the 5% milk/PBS-Tween buffer (0.1% v/v Tween 20 (Sigma P2287), 137mM NaCl (Sigma S9888), 2.7nM KCl (Sigma P4505), 10mM Na_2HPO_4 (Sigma S3264) and 2mM KH_2PO_4 (Sigma P9791), pH=7.4), and then overnight at 4-degrees Celsius with 1:5000 dilution of hORF1 (custom polyclonal rabbit antibody: TGNSKTQSASPPPK) antibody in the 3% milk/PBS-Tween buffer. Following the overnight incubation, the membrane was washed 3 times for 5 minutes a piece with

PBS-Tween buffer, and incubated for 1 hour at room temperature with 1:5000 dilution of horseradish peroxidase-conjugated secondary antibodies (HRP-donkey anti-rabbit (Santa Cruz: sc2317)) in 3% milk in PBS-Tween buffer. The membrane was washed one time for 15 minutes with PBS-Tween buffer then twice for 5 minutes with PBS-Tween buffer. The development was done using a 5-minute incubation with a Bio-Rad Clarity Kit (Bio-Rad). The signal was detected using a Chemi-Doc XRS+ Molecular Imager (Bio-Rad). anti-GAPDH antibodies (Santa Cruz: sc25778, 1:5000 dilution) were used to detect GAPDH using incubation conditions described above.

Figure S4.15C

Transfection: $7x10^5$ HeLa cells per T25 flask were seeded 16-18 hours prior to transfection. The cells were cotransfected with 1µg of a control plasmid (pBud), ORF1, e207, e127 or e259 plasmid and 1µg of either pBud or ORF1 plasmid using 4µl of Plus reagent in the total volume of 200µl of serum-free media. After 10 minutes of incubation at room temperature, 8µl of Lipofectamine mixed with 92µl of serum-free DMEM/High Glucose media was added to the reaction. The transfection cocktail was incubated for 15 minutes at room temperature and transferred into individual flasks containing 2mL of serum-free DMEM/High Glucose media. 3 hours post transfection, the media was replaced with 5mL of serum containing media. Cells were harvested approximately 24 hours later.

Protein harvest for western blot analysis: Cells were washed with 1X PBS (137mM NaCl (Sigma S9888), 2.7nM KCl (Sigma P4505), 10mM Na_2HPO_4 (Sigma S3264) and 2mM KH_2PO_4 (Sigma P9791), pH=7.4) and harvested using 300µl of TLB lysis buffer (50mM Tris, 150mM NaCl, 10mM EDTA, TritonX-100 0.5% v/v, Halt Protease inhibitor 10µl/mL, phosphatase inhibitors 2 and 3 (Sigma), pH=7.2) per T25 flask. The total lysate samples were sonicated five times for 5 seconds at 12 watts RMS using a 3mm wide Qsonica Microson homogenizer with Microson ultrasonic cell disruptor XL2000 (Microson). Samples were centrifuged at 21130xg for 15 minutes at 4-degrees Celsius. The resulting supernatant was transferred to a new microcentrifuge tube. The protein concentration for each sample was determined using 595nm wavelength OD values against a Bovine Serum Albumin (BSA) standard.

Western blot analysis performed under non-reducing conditions: 20µg of each sample was combined with 2x loading buffer to obtain the final concentration of 1X (4% SDS (Sigma L3771), 20% glycerol (v/v) (Sigma G5516), 0.004% bromophenol blue (Bio-Rad 161-0404), 0.125M Tris HCl (Sigma T3253), pH=6.8) and incubated at 85-degrees Celsius for 5 minutes prior to loading. Samples were fractionated on a Tris Acetate 3-8% Midi gel (Invitrogen) and transferred to a nitrocellulose membrane (iBlot2 system: Invitrogen). The membrane was first incubated for 1 hour in the 5% milk/PBS-Tween buffer (0.1% v/v Tween 20 (Sigma P2287), 137mM NaCl (Sigma S9888), 2.7nM KCl (Sigma P4505), 10mM Na_2HPO_4 (Sigma S3264) and 2mM KH_2PO_4 (Sigma

P9791), pH=7.4), and then overnight at 4-degrees Celsius with 1:5000 dilution of hORF1 (custom polyclonal rabbit antibody: TGNSKTQSASPPPK) antibody in the 3% milk/PBS-Tween buffer. Following the overnight incubation, the membrane was washed 3 times for 5 minutes a piece with PBS-Tween buffer, and incubated for 1 hour at room temperature with 1:5000 dilution of horseradish peroxidase-conjugated secondary antibodies (HRP-donkey anti-rabbit (Santa Cruz: sc2317)) in 3% milk in PBS-Tween buffer. The membrane was washed one time for 15 minutes with PBS-Tween buffer then twice for 5 minutes with PBS-Tween buffer. The development was done using a 5-minute incubation with a Bio-Rad Clarity Kit (Bio-Rad). The signal was detected using a Chemi-Doc XRS+ Molecular Imager (Bio-Rad). anti-GAPDH antibodies (Santa Cruz: sc25778, 1:5000 dilution) were used to detect GAPDH using incubation conditions described above.

Figure S4.16A

Generation of engineered HeLa cell lines harboring plasmids expressing L1wt or L1Stop: Same as in 8B, except each constitutively expressing HeLa flask was fed and passaged for additional 8 weeks.

Protein harvest for western blot analysis: $2x10^6$ HeLa cells were seeded into their own respective flask from each of the constitutively expressing HeLa cell lines. Flasks were harvested 40 hours after seeding. Cells were washed with 1X PBS. Cells were harvested using 500µl of TLB SDS (50mM Tris, 150mM NaCl, 10mM EDTA, 0.5% sodium dodecyl sulfate, TritonX-100 0.5% v/v, Halt Protease inhibitor 10µl/mL, phosphatase inhibitors 2 and 3 (Sigma), pH=7.2) per T75 flask. The total lysate samples were sonicated three times for 10 seconds at 12 watts RMS using a 3mm wide Qsonica Microson homogenizer with Microson ultrasonic cell disruptor XL2000 (Microson). Samples were centrifuged at 21130xg for 15 minutes at 4-degrees Celsius. The resulting supernatant was transferred to a new microcentrifuge tube. The protein concentration for each sample was determined using 595nm wavelength OD values against a Bovine Serum Albumin (BSA) standard.

Western blot analysis: 20µg of protein for each sample was combined with 2x Laemmli buffer to obtain the final concentration of 1X, 1.6µl β-mercaptoethanol and heated at 100-degrees Celsius for 5 minutes prior to loading. The positive control sample is the same sample from Figure 8B, hL1wt lane. Samples were fractionated on a Bis-Tris 4-12% Midi gel (Invitrogen) and transferred to a nitrocellulose membrane (iBlot2 system: Invitrogen). The membrane was first incubated for 1 hour in the 5% milk/PBS-Tween buffer (0.1% v/v Tween 20 (Sigma P2287), 137mM NaCl (Sigma S9888), 2.7nM KCl (Sigma P4505), 10mM Na_2HPO_4 (Sigma S3264) and 2mM KH_2PO_4 (Sigma P9791), pH=7.4), and then overnight at 4-degrees Celsius with 1:5000 dilution of hORF1 (custom polyclonal rabbit antibody: TGNSKTQSASPPPK) antibody in the 3% milk/PBS-Tween buffer. Following the overnight incubation, the membrane was washed 3 times for 5 minutes a piece with

PBS-Tween buffer, and incubated for 1 hour at room temperature with 1:5000 dilution of horseradish peroxidase-conjugated secondary antibodies (HRP-donkey anti-rabbit (Santa Cruz: sc2317)) in 3% milk in PBS-Tween buffer. The membrane was washed one time for 15 minutes with PBS-Tween buffer then twice for 5 minutes with PBS-Tween buffer. The development was done using a 5-minute incubation with a Bio-Rad Clarity Kit (Bio-Rad). The signal was detected using a Chemi-Doc XRS+ Molecular Imager (Bio-Rad). anti-GAPDH antibodies (Santa Cruz: sc25778, 1:5000 dilution) were used to detect GAPDH using incubation conditions described above.

Figure S4.16B

L1 retrotransposition assay: $5x10^5$ HeLa cells per T75 flask were seeded 16-18 hours prior to transfection. The cells were transfected with 0.1µg of L1Neo plasmid using 4µl of Plus reagent in the total volume of 200µl of serum-free media. After 10 minutes of incubation at room temperature, 8µl of Lipofectamine mixed with 92µl of serum-free DMEM/High Glucose media was added to the reaction. The transfection cocktail was incubated for 15 minutes at room temperature and transferred into individual flasks containing 6mL of serum-free DMEM/High Glucose media. 3 hours post transfection, the media was replaced with 10mL of serum containing media. 400µg/mL of G418 (Geneticin, Invitrogen: 10131-027) was administered 24 hours post transfection and maintained for up to 14 days with media changes every 2-3 days. The flasks were stained with 3mL of crystal violet staining solution (0.2% crystal violet (Sigma C6158), 5% acetic acid (Fisher Scientific A38-212), 2.5% isopropanol (Fisher Scientific BP2632-4)) per flask.

Transfection efficiency control: In the seeding with the L1 retrotransposition assay described above, a transfection efficiency control was done in parallel. $5x10^5$ HeLa cells per T75 flask were seeded 16-18 hours prior to transfection. The cells were transfected with 0.1µg of pIRES2-GFP plasmid using 4µl of Plus reagent in the total volume of 200µl of serum-free media. After 10 minutes of incubation at room temperature, 8µl of Lipofectamine mixed with 92µl of serum-free DMEM/High Glucose media was added to the reaction. The transfection cocktail was incubated for 15 minutes at room temperature and transferred into individual flasks containing 6mL of serum-free DMEM/High Glucose media. 3 hours post transfection, the media was replaced with 10mL of serum containing media. 400µg/mL of G418 (Geneticin, Invitrogen: 10131-027) was administered 24 hours post transfection and maintained for up to 14 days with media changes every 2-3 days. The flasks were stained with 3mL of crystal violet staining solution (0.2% crystal violet (Sigma C6158), 5% acetic acid (Fisher Scientific A38-212), 2.5% isopropanol (Fisher Scientific BP2632-4)) per flask.

References

1. N. A. Wallace, V. P. Belancio, P. L. Deininger, L1 mobile element expression causes multiple types of toxicity. *Gene* **419**, 75-81 (2008).
2. S. L. Gasior, T. P. Wakeman, B. Xu, P. L. Deininger, The Human LINE-1 Retrotransposon Creates DNA Double-strand Breaks. *Journal of Molecular Biology* **357**, 1383-1393 (2006).
3. E. Khazina *et al.*, Trimeric structure and flexibility of the L1ORF1 protein in human L1 retrotransposition. *Nat Struct Mol Biol* **18**, 1006-1014 (2011).
4. E. Khazina, O. Weichenrieder, Non-LTR retrotransposons encode noncanonical RRM domains in their first open reading frame. *Proceedings of the National Academy of Sciences* **106**, 731-736 (2009).
5. D. A. Kulpa, J. V. Moran, Ribonucleoprotein particle formation is necessary but not sufficient for LINE-1 retrotransposition. *Human Molecular Genetics* **14**, 3237-3248 (2005).
6. T. Penzkofer, T. Dandekar, T. Zemojtel, L1Base: from functional annotation to prediction of active LINE-1 elements. *Nucleic Acids Research* **33**, D498-D500 (2005).
7. J. V. Moran *et al.*, High Frequency Retrotransposition in Cultured Mammalian Cells. *Cell* **87**, 917-927 (1996).
8. P. Deininger *et al.* (NAR, 2016).
9. M. Sokolowski, D. deHaro, C. M. Christian, K. J. Kines, V. P. Belancio, Characterization of L1 ORF1p Self-Interaction and Cellular Localization Using a Mammalian Two-Hybrid System. *PLoS ONE* **8**, e82021 (2013).
10. Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921 (2001).
11. R. Cordaux, M. A. Batzer, The impact of retrotransposons on human genome evolution. *Nat Rev Genet* **10**, 691-703 (2009).
12. S. L. Martin *et al.*, A single amino acid substitution in ORF1 dramatically decreases L1 retrotransposition and provides insight into nucleic acid chaperone activity. *Nucleic Acids Research* **36**, 5845-5854 (2008).
13. M. Dewannieux, C. Esnault, T. Heidmann, LINE-mediated retrotransposition of marked Alu sequences. *Nat Genet* **35**, (2003).
14. E. M. Ostertag, J. L. Goodier, Y. Zhang, H. H. Kazazian Jr, SVA Elements Are Nonautonomous Retrotransposons that Cause Disease in Humans. *The American Journal of Human Genetics* **73**, 1444-1451 (2003).
15. N. Wallace, B. J. Wagstaff, P. L. Deininger, A. M. Roy-Engel, LINE-1 ORF1 protein enhances Alu SINE retrotransposition. *Gene* **419**, 1-6 (2008).
16. S. L. Martin *et al.*, LINE-1 Retrotransposition Requires the Nucleic Acid Chaperone Activity of the ORF1 Protein. *Journal of Molecular Biology* **348**, 549-561 (2005).
17. J. D. Evans, S. Peddigari, K. R. Chaurasiya, M. C. Williams, S. L. Martin, Paired mutations abolish and restore the balanced annealing and melting activities of ORF1p that are required for LINE-1 retrotransposition. *Nucleic Acids Research* **39**, 5611-5621 (2011).
18. C. Esnault, J. Maestre, T. Heidmann, Human LINE retrotransposons generate processed pseudogenes. *Nat Genet* **24**, (2000).
19. H. H. Kazazian *et al.*, Haemophilia A resulting from de novo insertion of L1 sequences represents a novel mechanism for mutation in man. *Nature* **332**, 164-166 (1988).
20. P. Woods-Samuels *et al.*, Characterization of a nondeleterious L1 insertion in an intron of the human factor VIII gene and further evidence of open reading frames in functional L1 elements. *Genomics* **4**, 290-296 (1989).

21. R. S. Alisch, J. L. Garcia-Perez, A. R. Muotri, F. H. Gage, J. V. Moran, Unconventional translation of mammalian LINE-1 retrotransposons. *Genes & Development* **20**, 210-224 (2006).
22. B. Brouha *et al.*, Hot L1s account for the bulk of retrotransposition in the human population. *Proc Natl Acad Sci USA* **100**, (2003).
23. C. R. Beck, J. L. Garcia-Perez, R. M. Badge, J. V. Moran, LINE-1 Elements in Structural Variation and Disease. *Annual Review of Genomics and Human Genetics* **12**, 187-215 (2011).
24. C. R. L. Huang *et al.*, Mobile Interspersed Repeats Are Major Structural Variants in the Human Genome. *Cell* **141**, 1171-1182 (2010).
25. A. D. Ewing, H. H. Kazazian, Whole-genome resequencing allows detection of many rare LINE-1 insertion alleles in humans. *Genome Research* **21**, 985-990 (2011).
26. C. R. Beck *et al.*, LINE-1 Retrotransposition Activity in Human Genomes. *Cell* **141**, 1159-1170 (2010).
27. M. K. Konkel, J. Wang, P. Liang, M. A. Batzer, Identification and characterization of novel polymorphic LINE-1 insertions through comparison of two human genome sequence assemblies. *Gene* **390**, 28-38 (2007).
28. J. V. Moran, High frequency retroposition in cultured mammalian cells. *Cell* **87**, 917-927 (1996).
29. V. O. Kolosha, S. L. Martin, In vitro properties of the first ORF protein from mouse LINE-1 support its role in ribonucleoprotein particle formation during retrotransposition. *Proceedings of the National Academy of Sciences* **94**, 10155-10160 (1997).
30. D. A. Kulpa, J. V. Moran, Cis-preferential LINE-1 reverse transcriptase activity in ribonucleoprotein particles. *Nat Struct Mol Biol* **13**, 655-660 (2006).
31. Q. Feng, J. V. Moran, H. H. Kazazian, J. D. Boeke, Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. *Cell* **87**, (1996).
32. Y. Xiong, T. H. Eickbush, Origin and evolution of retroelements based upon their reverse transcriptase sequences. *The EMBO Journal* **9**, 3353-3362 (1990).
33. S. L. Mathias, A. F. Scott, H. H. Kazazian, J. D. Boeke, A. Gabriel, Reverse transcriptase encoded by a human transposable element. *Science* **254**, 1808-1810 (1991).
34. D. D. Luan, M. H. Korman, J. L. Jakubczak, 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* **72**, (1993).
35. G. J. Cost, Q. Feng, A. Jacquier, J. D. Boeke, Human L1 element target-primed reverse transcription in vitro. *EMBO J* **21**, (2002).
36. W. Wei *et al.*, Human L1 Retrotransposition: cisPreference versus trans Complementation. *Molecular and Cellular Biology* **21**, 1429-1439 (2001).
37. K. E. Callahan, A. B. Hickman, C. E. Jones, R. Ghirlando, A. V. Furano, Polymerization and nucleic acid-binding properties of human L1 ORF1 protein. *Nucleic Acids Research* **40**, 813-827 (2012).
38. S. L. Martin, Nucleic acid chaperone properties of ORF1p from the non-LTR retrotransposon, LINE-1. *RNA Biology* **7**, 706-711 (2010).
39. S. L. Martin, D. Branciforte, D. Keller, D. L. Bain, Trimeric structure for an essential protein in L1 retrotransposition. *Proceedings of the National Academy of Sciences* **100**, 13815-13820 (2003).
40. P. R. Cook, C. E. Jones, A. V. Furano, Phosphorylation of ORF1p is required for L1 retrotransposition. *Proceedings of the National Academy of Sciences* **112**, 4298-4303 (2015).

41. S. Basame *et al.*, Spatial Assembly and RNA Binding Stoichiometry of a LINE-1 Protein Essential for Retrotransposition. *Journal of Molecular Biology* **357**, 351-357 (2006).
42. M. N. Naufer *et al.*, L1 retrotransposition requires rapid ORF1p oligomerization, a novel coiled coil-dependent property conserved despite extensive remodeling. *Nucleic Acids Research* **44**, 281-293 (2016).
43. S. L. Martin, J. Li, J. A. Weisz, Deletion analysis defines distinct functional domains for protein-protein and nucleic acid interactions in the ORF1 protein of mouse LINE-11. *Journal of Molecular Biology* **304**, 11-20 (2000).
44. S. Boissinot, A. V. Furano, Adaptive Evolution in LINE-1 Retrotransposons. *Molecular Biology and Evolution* **18**, 2186-2194 (2001).
45. S. Boissinot, A. Sookdeo, The Evolution of LINE-1 in Vertebrates. *Genome Biology and Evolution* **8**, 3485-3507 (2016).
46. R. A. Kammerer *et al.*, A conserved trimerization motif controls the topology of short coiled coils. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 13891-13896 (2005).
47. J. L. Goodier, L. Zhang, M. R. Vetter, H. H. Kazazian, 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**, (2007).
48. J. L. Goodier, L. E. Cheung, H. H. Kazazian, Mapping the LINE1 ORF1 protein interactome reveals associated inhibitors of human retrotransposition. *Nucleic Acids Research*, (2013).
49. Martin S. Taylor *et al.*, Affinity Proteomics Reveals Human Host Factors Implicated in Discrete Stages of LINE-1 Retrotransposition. *Cell* **155**, 1034-1048 (2013).
50. S. Kubo *et al.*, L1 retrotransposition in nondividing and primary human somatic cells. *Proceedings of the National Academy of Sciences* **103**, 8036-8041 (2006).
51. Y. Xie *et al.*, Cell division promotes efficient retrotransposition in a stable L1 reporter cell line. *Mobile DNA* **4**, 10 (2013).
52. J. L. Goodier, E. M. Ostertag, K. A. Engleka, M. C. Seleme, H. H. Kazazian, A potential role for the nucleolus in L1 retrotransposition. *Hum Mol Genet* **13**, (2004).
53. S. Ergun *et al.*, Cell type-specific expression of LINE-1 open reading frames 1 and 2 in fetal and adult human tissues. *J Biol Chem* **279**, (2004).
54. L. Chen, J. Dahlstrom, A. Chandra, P. Board, D. Rangasamy, Prognostic value of LINE-1 retrotransposon expression and its subcellular localization in breast cancer. *Breast Cancer Research and Treatment* **136**, 129-142 (2012).
55. C. R. Harris *et al.*, Association of Nuclear Localization of a Long Interspersed Nuclear Element-1 Protein in Breast Tumors with Poor Prognostic Outcomes. *Genes & Cancer* **1**, 115-124 (2010).
56. N. Rodić *et al.*, Long Interspersed Element-1 Protein Expression Is a Hallmark of Many Human Cancers. *The American Journal of Pathology* **184**, 1280-1286 (2014).
57. T. T. Doucet-O'Hare *et al.*, Somatically Acquired LINE-1 Insertions in Normal Esophagus Undergo Clonal Expansion in Esophageal Squamous Cell Carcinoma. *Human Mutation* **37**, 942-954 (2016).
58. T. T. Doucet-O'Hare *et al.*, LINE-1 expression and retrotransposition in Barrett's esophagus and esophageal carcinoma. *Proceedings of the National Academy of Sciences* **112**, E4894-E4900 (2015).
59. A. Sookdeo, C. M. Hepp, M. A. McClure, S. Boissinot, Revisiting the evolution of mouse LINE-1 in the genomic era. *Mobile DNA* **4**, 1-15 (2013).

60. K. J. Kines, M. Sokolowski, D. L. deHaro, C. M. Christian, V. P. Belancio, Potential for genomic instability associated with retrotranspositionally-incompetent L1 loci. *Nucleic Acids Research*, (2014).
61. S. H. Rangwala, L. Zhang, H. H. Kazazian, Many LINE1 elements contribute to the transcriptome of human somatic cells. *Genome Biology* **10**, R100 (2009).
62. M. Dewannieux, C. Esnault, T. Heidmann, LINE-mediated retrotransposition of marked Alu sequences. *Nat Genet* **35**, 41-48 (2003).
63. C. Esnault, J. Maestre, T. Heidmann, Human LINE retrotransposons generate processed pseudogenes. *Nat Genet* **24**, 363-367 (2000).
64. D. D. Luan, M. H. Korman, J. L. Jakubczak, 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* **72**, 595-605 (1993).
65. G. J. F. Q. J. A. B. J. D. Cost, Human L1 element target-primed reverse transcription in vitro. *The EMBO Journal* **21**, 5899-5910 (2002).
66. B. Brouha *et al.*, Hot L1s account for the bulk of retrotransposition in the human population. *Proceedings of the National Academy of Sciences* **100**, 5280-5285 (2003).
67. Q. Feng, J. V. Moran, H. H. Kazazian Jr, J. D. Boeke, Human L1 Retrotransposon Encodes a Conserved Endonuclease Required for Retrotransposition. *Cell* **87**, 905-916 (1996).
68. Y. Xiong, T. H. Eickbush, Origin and evolution of retroelements based upon their reverse transcriptase sequences. *The EMBO journal* **9**, 3353 (1990).
69. T. G. Fanning, M. F. Singer, LINE-1: A mammalian transposable element. *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression* **910**, 203-212 (1987).
70. O. Piskareva, C. Ernst, N. Higgins, V. Schmatchenko, The carboxy-terminal segment of the human LINE-1 ORF2 protein is involved in RNA binding. *FEBS Open Bio* **3**, 433-437 (2013).
71. P. W.-L. Li, J. Li, S. L. Timmerman, L. A. Krushel, 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 Research* **34**, 853-864 (2006).
72. S. Ergün *et al.*, Cell Type-specific Expression of LINE-1 Open Reading Frames 1 and 2 in Fetal and Adult Human Tissues. *Journal of Biological Chemistry* **279**, 27753-27763 (2004).
73. J. L. Goodier, E. M. Ostertag, K. A. Engleka, M. C. Seleme, H. H. Kazazian, A potential role for the nucleolus in L1 retrotransposition. *Human Molecular Genetics* **13**, 1041-1048 (2004).
74. A. J. Doucet *et al.*, Characterization of LINE-1 Ribonucleoprotein Particles. *PLoS Genet* **6**, e1001150 (2010).
75. K. Repanas *et al.*, Determinants for DNA target structure selectivity of the human LINE-1 retrotransposon endonuclease. *Nucleic Acids Research* **35**, 4914-4926 (2007).
76. I. V. Nesterova *et al.*, Near-IR single fluorophore quenching system based on phthalocyanine (Pc) aggregation and its application for monitoring inhibitor/activator action on a therapeutic target: L1-EN. *Analyst* **136**, 1103-1105 (2011).
77. V. Perepelitsa-Belancio, P. Deininger, RNA truncation by premature polyadenylation attenuates human mobile element activity. *Nat Genet* **35**, 363-366 (2003).
78. B. J. Wagstaff, M. Barnerβoi, A. M. Roy-Engel, Evolutionary Conservation of the Functional Modularity of Primate and Murine LINE-1 Elements. *PLoS ONE* **6**, e19672 (2011).

79. O. Weichenrieder, K. Repanas, A. Perrakis, Crystal Structure of the Targeting Endonuclease of the Human LINE-1 Retrotransposon. *Structure* **12**, 975-986 (2004).
80. S. Madhusudan *et al.*, Isolation of a small molecule inhibitor of DNA base excision repair. *Nucleic Acids Research* **33**, 4711-4724 (2005).
81. H. Khan, A. Smit, S. Boissinot, Molecular evolution and tempo of amplification of human LINE-1 retrotransposons since the origin of primates. *Genome Research* **16**, 78-87 (2006).
82. T. P. Naas *et al.*, An actively retrotransposing, novel subfamily of mouse L1 elements. *EMBO J* **17**, 590-597 (1998).
83. T. A. Morrish *et al.*, DNA repair mediated by endonuclease-independent LINE-1 retrotransposition. *Nat Genet* **31**, 159-165 (2002).
84. J. S. Han, J. D. Boeke, A highly active synthetic mammalian retrotransposon. *Nature* **429**, 314-318 (2004).
85. C. Robberecht, T. Voet, M. Z. Esteki, B. A. Nowakowska, J. R. Vermeesch, Nonallelic homologous recombination between retrotransposable elements is a driver of de novo unbalanced translocations. *Genome Research* **23**, 411-418 (2013).
86. B. A. Dombroski, S. L. Mathias, E. Nanthakumar, A. F. Scott, H. H. Kazazian, Isolation of an active human transposable element. *Science* **254**, 1805-1808 (1991).
87. K. Han *et al.*, L1 recombination-associated deletions generate human genomic variation. *Proceedings of the National Academy of Sciences* **105**, 19366-19371 (2008).
88. Y. Miki *et al.*, Disruption of the APC Gene by a Retrotransposal Insertion of L1 Sequence in a Colon Cancer. *Cancer Research* **52**, 643-645 (1992).
89. O. K. Pickeral, W. Makałowski, M. S. Boguski, J. D. Boeke, Frequent Human Genomic DNA Transduction Driven by LINE-1 Retrotransposition. *Genome Research* **10**, 411-415 (2000).
90. A. Zhang *et al.*, RNase L restricts the mobility of engineered retrotransposons in cultured human cells. *Nucleic Acids Research* **42**, 3803-3820 (2014).
91. N. Yang, H. H. Kazazian, L1 retrotransposition is suppressed by endogenously encoded small interfering RNAs in human cultured cells. *Nat Struct Mol Biol* **13**, 763-771 (2006).
92. A. V. Horn *et al.*, Human LINE-1 restriction by APOBEC3C is deaminase independent and mediated by an ORF1p interaction that affects LINE reverse transcriptase activity. *Nucleic Acids Research* **42**, 396-416 (2014).
93. V. P. Belancio, D. J. Hedges, P. Deininger, LINE-1 RNA splicing and influences on mammalian gene expression. *Nucleic Acids Research* **34**, 1512-1521 (2006).
94. H. P. Bogerd *et al.*, Cellular inhibitors of long interspersed element 1 and Alu retrotransposition. *Proceedings of the National Academy of Sciences* **103**, 8780-8785 (2006).
95. D. deHaro *et al.*, 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**, 7694-7707 (2014).
96. L. Dai, Q. Huang, J. Boeke, Effect of reverse transcriptase inhibitors on LINE-1 and Ty1 reverse transcriptase activities and on LINE-1 retrotransposition. *BMC Biochemistry* **12**, 18 (2011).
97. E. N. Kroutter, V. P. Belancio, B. J. Wagstaff, A. M. Roy-Engel, The RNA Polymerase Dictates ORF1 Requirement and Timing of LINE and SINE Retrotransposition. *PLoS Genet* **5**, e1000458 (2009).

98. C. Strahl, E. H. Blackburn, Effects of reverse transcriptase inhibitors on telomere length and telomerase activity in two immortalized human cell lines. *Molecular and Cellular Biology* **16**, 53-65 (1996).
99. G. D'Andrea, F. Brisdelli, A. Bozzi, AZT: an old drug with new perspectives. *Curr Clin Pharmacol* **3**, 20-37 (2008).
100. K. Imai, A. Takaoka, Comparing antibody and small-molecule therapies for cancer. *Nat Rev Cancer* **6**, 714-727 (2006).
101. M. Hidalgo, S. G. Eckhardt, Development of Matrix Metalloproteinase Inhibitors in Cancer Therapy. *Journal of the National Cancer Institute* **93**, 178-193 (2001).
102. S. L. Martin, F. D. Bushman, Nucleic Acid Chaperone Activity of the ORF1 Protein from the Mouse LINE-1 Retrotransposon. *Molecular and Cellular Biology* **21**, 467-475 (2001).
103. J. L. Goodier, L. Zhang, M. R. Vetter, H. H. Kazazian, LINE-1 ORF1 Protein Localizes in Stress Granules with Other RNA-Binding Proteins, Including Components of RNA Interference RNA-Induced Silencing Complex. *Molecular and Cellular Biology* **27**, 6469-6483 (2007).
104. J. L. Goodier, L. E. Cheung, H. H. Kazazian, Jr., MOV10 RNA Helicase Is a Potent Inhibitor of Retrotransposition in Cells. *PLoS Genet* **8**, e1002941 (2012).
105. V. P. Belancio, Importance of RNA analysis in interpretation of reporter gene expression data. *Analytical Biochemistry* **417**, 159-161 (2011).
106. S. E. Dmitriev *et al.*, Efficient Translation Initiation Directed by the 900-Nucleotide-Long and GC-Rich 5' Untranslated Region of the Human Retrotransposon LINE-1 mRNA Is Strictly Cap Dependent Rather than Internal Ribosome Entry Site Mediated. *Molecular and Cellular Biology* **27**, 4685-4697 (2007).
107. D. Fass, Disulfide Bonding in Protein Biophysics. *Annual Review of Biophysics* **41**, 63-79 (2012).
108. T. P. Dick, N. Bangia, D. R. Peaper, P. Cresswell, Disulfide Bond Isomerization and the Assembly of MHC Class I-Peptide Complexes. *Immunity* **16**, 87-98 (2002).
109. M. Molinari, A. Helenius, Glycoproteins form mixed disulphides with oxidoreductases during folding in living cells. *Nature* **402**, 90-93 (1999).
110. V. P. Belancio, D. J. Hedges, P. Deininger, Mammalian non-LTR retrotransposons: For better or worse, in sickness and in health. *Genome Research* **18**, 343-358 (2008).
111. D. D. Luan, T. H. Eickbush, Downstream 28S gene sequences on the RNA template affect the choice of primer and the accuracy of initiation by the R2 reverse transcriptase. *Molecular and Cellular Biology* **16**, 4726-4734 (1996).
112. S. L. Gasior *et al.*, Characterization of pre-insertion loci of de novo L1 insertions. *Gene* **390**, 190-198 (2007).
113. J. M. Tavare, L. M. Fletcher, G. I. Welsh, Using green fluorescent protein to study intracellular signalling. *Journal of Endocrinology* **170**, 297-306 (2001).
114. S. Malki, Godfried W. van der Heijden, Kathryn A. O'Donnell, Sandra L. Martin, A. Bortvin, A Role for Retrotransposon LINE-1 in Fetal Oocyte Attrition in Mice. *Developmental Cell* **29**, 521-533 (2014).
115. S. F. C. Soper *et al.*, Mouse Maelstrom, a Component of Nuage, Is Essential for Spermatogenesis and Transposon Repression in Meiosis. *Developmental Cell* **15**, 285-297 (2008).
116. B. J. Wagstaff, E. N. Kroutter, R. S. Derbes, V. P. Belancio, A. M. Roy-Engel, Molecular Reconstruction of Extinct LINE-1 Elements and Their Interaction with Nonautonomous Elements. *Molecular Biology and Evolution* **30**, 88-99 (2013).

117. K. J. Kines, M. Sokolowski, D. L. deHaro, C. M. Christian, V. P. Belancio, Potential for genomic instability associated with retrotranspositionally-incompetent L1 loci. *Nucleic Acids Research* **42**, 10488-10502 (2014).
118. S. Rangwala, L. Zhang, H. Kazazian, Many LINE1 elements contribute to the transcriptome of human somatic cells. *Genome Biology* **10**, R100 (2009).
119. D. M. Sassaman *et al.*, Many human L1 elements are capable of retrotransposition. *Nat Genet* **16**, 37-43 (1997).
120. K. J. Kines *et al.*, The endonuclease domain of the LINE-1 ORF2 protein can tolerate multiple mutations. *Mobile DNA* **7**, 8 (2016).
121. M. Sokolowski *et al.*, Development of a monoclonal antibody specific to the endonuclease domain of the human LINE-1 ORF2 protein. *Mobile DNA* **5**, 29 (2014).
122. C. M. Christian, D. deHaro, K. J. Kines, M. Sokolowski, V. P. Belancio, Identification of L1 ORF2p sequence important to retrotransposition using Bipartile Alu retrotransposition (BAR). *Nucleic Acids Research* **44**, 4818-4834 (2016).
123. C. M. Christian, M. Sokolowski, D. deHaro, K. J. Kines, V. P. Belancio, Involvement of Conserved Amino Acids in the C-Terminal Region of LINE-1 ORF2p in Retrotransposition. *Genetics*, (2017).
124. H. K. Surks *et al.*, Regulation of Myosin Phosphatase by a Specific Interaction with cGMP-Dependent Protein Kinase Iα. *Science* **286**, 1583-1587 (1999).
125. S. S. Chen, C. N. Lee, W. R. Lee, K. McIntosh, T. H. Lee, Mutational analysis of the leucine zipper-like motif of the human immunodeficiency virus type 1 envelope transmembrane glycoprotein. *Journal of Virology* **67**, 3615-3619 (1993).
126. H. Y. Cheng, A. P. Schiavone, T. E. Smithgall, A Point Mutation in the N-Terminal Coiled-Coil Domain Releases c-Fes Tyrosine Kinase Activity and Survival Signaling in Myeloid Leukemia Cells. *Molecular and Cellular Biology* **21**, 6170-6180 (2001).
127. M. Speek, Antisense Promoter of Human L1 Retrotransposon Drives Transcription of Adjacent Cellular Genes. *Molecular and Cellular Biology* **21**, 1973-1985 (2001).
128. Ahmet M. Denli *et al.*, Primate-Specific ORF0 Contributes to Retrotransposon-Mediated Diversity. *Cell* **163**, 583-593 (2015).
129. V. P. Belancio, A. M. Roy-Engel, P. Deininger, The impact of multiple splice sites in human L1 elements. *Gene* **411**, 38-45 (2008).
130. P. Deininger, V. P. Belancio, in *Transposons and Retrotransposons: Methods and Protocols,* L. J. Garcia-Pérez, Ed. (Springer New York, New York, NY, 2016), pp. 223-236.
131. N. Rodic *et al.*, Retrotransposon insertions in the clonal evolution of pancreatic ductal adenocarcinoma. *Nat Med* **21**, 1060-1064 (2015).
132. C. Philippe *et al.*, Activation of individual L1 retrotransposon instances is restricted to cell-type dependent permissive loci. *eLife* **5**, e13926 (2016).
133. J. L. Goodier, G. C. Pereira, L. E. Cheung, R. J. Rose, H. H. Kazazian, Jr., The Broad-Spectrum Antiviral Protein ZAP Restricts Human Retrotransposition. *PLoS Genet* **11**, e1005252 (2015).
134. J. B. Moldovan, J. V. Moran, The Zinc-Finger Antiviral Protein ZAP Inhibits LINE and Alu Retrotransposition. *PLoS Genet* **11**, e1005121 (2015).
135. R. N. McLaughlin, J. T. Gable, C. J. Wittkopp, M. Emerman, H. S. Malik, Conservation and Innovation of APOBEC3A Restriction Functions during Primate Evolution. *Molecular Biology and Evolution*, (2016).
136. G. Servant *et al.*, The Nucleotide Excision Repair Pathway Limits L1 Retrotransposition. *Genetics* **205**, 139 (2017).

137. V. P. Belancio, LINE-1 activity as molecular basis for genomic instability associated with light exposure at night. *Mobile Genetic Elements* **5**, 46-50 (2015).
138. W. An *et al.*, Characterization of a synthetic human LINE-1 retrotransposon ORFeus-Hs. *Mobile DNA* **2**, 2 (2011).
139. A. F. A. Smit, Interspersed repeats and other mementos of transposable elements in mammalian genomes. *Current Opinion in Genetics & Development* **9**, 657-663 (1999).
140. J. Jurka *et al.*, Repbase Update, a database of eukaryotic repetitive elements. *Cytogenetic and Genome Research* **110**, 462-467 (2005).
141. W. Bao, J. Jurka, Origin and evolution of LINE-1 derived "half-L1" retrotransposons (HAL1). *Gene* **465**, 9-16 (2010).
142. V. P. Belancio, D. J. Hedges, P. Deininger, LINE-1 RNA splicing and influences on mammalian gene expression. *Nucleic Acids Res* **34**, (2006).
143. P. Deininger *et al.*, A comprehensive approach to expression of L1 loci. *Nucleic Acids Research* **45**, e31-e31 (2017).
144. V. P. Belancio, A. M. Roy-Engel, R. R. Pochampally, P. Deininger, Somatic expression of LINE-1 elements in human tissues. *Nucleic Acids Research* **38**, 3909-3922 (2010).
145. A. F. A. Smit, G. Tóth, A. D. Riggs, J. Jurka, Ancestral, Mammalian-wide Subfamilies of LINE-1 Repetitive Sequences. *Journal of Molecular Biology* **246**, 401-417 (1995).
146. P. L. Deininger, M. A. Batzer, C. A. Hutchison, M. H. Edgell, Master genes in mammalian repetitive DNA amplification. *Trends in Genetics* **8**, 307-311 (1992).
147. M. Mears, C. Hutchison, III, The Evolution of Modern Lineages of Mouse L1 Elements. *Journal of Molecular Evolution* **52**, 51-62 (2001).
148. P. Wincker, V. Jubier-Maurin, G. Roizés, Unrelated sequences at the 5' end of mouse LINE-1 repeated elements define two distinct sub-families. *Nucleic Acids Research* **15**, 8593-8606 (1987).
149. R. J. DeBerardinis, J. L. Goodier, E. M. Ostertag, H. H. Kazazian, Rapid amplification of a retrotransposon subfamily is evolving the mouse genome. *Nat Genet* **20**, 288-290 (1998).
150. S. C. Hardies *et al.*, LINE-1 (L1) Lineages in the Mouse. *Molecular Biology and Evolution* **17**, 616-628 (2000).

Biography

Mark Sokolowski attended Loyola University Chicago for his Bachelor of Science (B.S.) in Biology and Bachelor of Arts (B.A.) in Spanish, and graduated in May of 2010. He attended Tulane University's School of Medicine BMS Graduate Program in August of 2010. Shortly after, he joined Dr. Victoria P. Belancio's laboratory in spring of 2011 to pursue his Ph.D. study of the LINE-1 retrotransposon. He will be joining Dr. David Largaespada's laboratory at the Masonic Cancer Center at the University of Minnesota as a postdoc to study cancer biology.