COMPREHENSIVE SCANNING MUTAGENESIS OF A HUMAN RETROTRANSPOSON IDENTIFIES MOTIFS ESSENTIAL FOR FUNCTION

by
Emily Adney

A dissertation submitted to Johns Hopkins University in conformity
with the requirements for the degree of Doctor of Philosophy

Baltimore, Maryland
December 2018

Abstract

Long Interspersed Nuclear Element-1 (LINE-1, L1) is the only autonomous active transposable element in the human genome. In general, we strive for molecular level understanding of how the L1- encoded proteins ORF1p and ORF2p facilitate retrotransposition as they are essential for enabling this element to jump from one locus to another via a "copy and paste" mechanism. In this work, we aimed to develop a variety of tools to probe specific intermolecular interactions that form with RNA, proteins and target DNA throughout the L1 lifecycle. ORF1p is an RNA-binding protein and ORF2p has both endonuclease and reverse transcriptase activities. These proteins bind the L1 RNA to form L1 ribonucleoprotein complexes (RNPs). As a streamlined parasite, the L1 retrotransposon requires a variety of host factors to complete a successful lifecycle and the host has likely mainly evolved to limit the mutagenic potential of novel L1 insertions. First, we study L1 RNP formation *in vivo* and, second, we study the L1 encoded proteins' sensitivity to mutation. As a follow-up to studying the composition of RNPs in tissue culture, we established customized tools for isolating active L1 RNP complexes from live mammalian tissues. This necessitated the establishment of novel transgenic mouse lines highly expressing tagged mouse L1 proteins, as well as the production of a high-quality antibody against mouse ORF1p. Prior studies used human L1 in mouse, and thus these studies represent a truly homologous system. We also successfully conducted a mutagenic scan of human L1 by constructing a library consisting of 538 consecutive trialanine substitutions, scanning along ORF1p and ORF2p. We describe the construction of the library, its initial characterization, and its use as a resource for future studies. For each variant, we measured retrotransposition efficiency. We also measured both total ORF1p and RNA produced by each variant. We also started to develop an RNA sequencing-based method to quantify how well each ORF1 variant protein was able to bind its own L1 RNA for proper RNP formation. Retrotransposition was extremely sensitive to mutations in ORF1p and ORF2p. The library provides comprehensive information on which regions are most critical to retrotransposition and which are dispensable.

Jef D. Boeke, Ph.D. (Sponsor & Reader)
Professor
Institute for Systems Genetics
NYU Langone Health

Kathleen H. Burns, M.D., Ph.D. (Reader)
Professor
Department of Pathology
McKusick-Nathans Institute of Genetic Medicine
Johns Hopkins University School of Medicine

Acknowledgements

Although on *page iv*, I write this last, as I prepare to hit the "submit thesis" button and become a doctor! Wow! I am not ashamed to say that I have tried to write this section as fully and whole-heartedly as possible many times now. Every time, I get choked up and tears well up… (surprised?)

I have tried to detail my due acknowledgement for everyone involved in my graduate school experience, which includes a Masters in Molecular Biophysics in the laboratory of Juliette Lecomte (data not shown) and a PhD in Human Genetics with Jef Boeke, both from Johns Hopkins. (Johns Hopkins!? What a dream! Two programs? Even bigger dream!) This all may have been an inducer of some "imposter syndrome", but mainly has provided me with the most stellar education, outstanding scientific community, and brilliant confidence to continue in academia… with great success, we can only hope. These experiences have well-surpassed what I could have ever dreamed of. I am honored that this thesis will live in the Hopkins archives. Thank you, Johns Hopkins!

The most prominent feeling I have as I finish graduate school is gratitude for all the people involved -- mentors, friends, family, colleagues -- I have been so extremely lucky to be surrounded by such smart, talented, generous, fun, passionate, and loving people. You can't do anything alone… and I never felt like I had to. So, thank you all.

That being said, I would like to write a few brief personalized notes to the following people….

Jef Boeke

You have been an incredible mentor. Thank you for immersing your lab in an environment stimulated by a genuine passion for every detail of science. Thank you for mentoring me and being so caring for me as a person as a whole. You have taught me so much in science, life and in the art of moving labs across state lines, and within a three-block radius approximately five times. I am impressed with the institute that you direct. You have put so much work into nurturing and expanding it for the last few years and it flourishes. It is an absolutely incredible place to work. Thank you. You are a wonderful boss and I don't think I will ever stop coming to you for great scientific chats. I am so honored to have been invited to join you for my PhD-earning journey. You have helped prepare me well (from what I can tell) for the science I do in the future. Cheers!

Kathy Burns

You are a fantastic role model and friend and I aspire to be like you. Your scientific expertise has helped me so much. Thank you for being my formal *thesis reader*, thesis committee member, and provider of recommendations. You are a fabulous person and your impact on me has been profound in such a radiantly positive way.

David Fenyö and Liam Holt

Thank you both for being wonderful colleagues, collaborators, mentors, friends, and the most fun hiking and yoga companions. I will miss popping into your offices and saying hello…. Just kidding I am going to keep doing that forever. Your energy in ISG are two in a million and I appreciate you both for it.

Thesis Committee

Jeremy Nathans, David Graham, Roger Reeves, and Kathy Burns – thank you for being supportive and wonderful guides.

Your wisdom and perspective made my work better in so many ways.

David Valle, Sandy Muscelli, and the Human Genetics Program

Thank you for accepting me and training me! What an incredible environment you have created in the graduate training program -- one that I am so thrilled to be a part of. I appreciate all of your hard work and talent. I will never forget the course in Bar Harbor, it changed my life. Thank you for being so supportive… especially as I freaked out before the Graduate Board Oral exam, as I skyped into Journal Clubs, and as I made day trips from New York to complete my training with you.

Juliette Lecomte, David Shortle, Ranice Crosby, and Ananya Majumdar

Thank you for the mentorship and support during my training in biophysics and afterwards.

Boeke Labbies : Past and Present

You have been like a solid family and you all rock so much!

Special thanks to those who helped train me :

Neta Agmon, Paolo Mita, David Truong, Leslie Mitchell,

Marty Taylor, Lixin Dai, and John LaCava

Special thanks to the other L1 team members!

Special thanks to those who helped with administrative work :

Julie Oaks, Nakisha Davis, Jessica Lorenzo, Teddy Shin, Andrew Martin and Deb Bemis

Thank you to all others for making lab a place I looked forward to being daily.

More friends

Carla, Julie, Lisa, Neta, Donghui, Aleks, Jon, Jasmine, Sarah, Sud, Andrew, and Anne…. How did I get so lucky!? Thank you for spiking in so much love and fun to the time spent both in and out of lab. You are the some of the most awesome people on the planet.

Family

…. back to that choked up conundrum.

Thus, I simply say thank you and I love you so very much…

Mom

Dad

Ralph

Brandyn

Alma

Ali

Omar

Grandpa Brown

…

Table of Contents

Title.. i

Abstract.. ii

Acknowledgements.. v

List of Figures ix

List of Tables xi

Chapter 1 : Introduction to L1 retrotransposition...................................... 1

Chapter 2 : Tool development for studying L1 RNPs formed *in vivo*................. 10

Summary.. 10

Introduction... 10

Results and Discussion

- Production of anti-mouse ORF1 antibodies.. 13
- Establishment of tagged L1 transgenic mouse lines.. 18
- Initial tests of targeted isolation of protein from cryo-milled mouse tissue 29

Methods 34

Chapter 3 : Comprehensive scanning mutagenesis of the human retrotransposon L1 identifies novel motifs essential for function.............................. 39

Summary.. 39

Introduction... 40

Results and Discussion

- Designing and building a trialanine scanning mutagenic library.............................. 42
- Measurements of the retrotransposition frequency of each mutant............................. 54
- Quantification of ORF1p variant protein abundance 69
- Probing which ORF1 mutants form RNPs with L1 RNA : pooled screening....... 75
- Investigation of trends in nucleolar localization of certain ORF1p trialanine mutants.. 91
- Trends in amino acid conservation and sensitivity to mutation across ORF2p.... 98

Methods 112

References 132

Curriculum Vitae 140

List of Figures

Figure 1.1 : The structure and lifecycle of L1.. 5

Figure 1.2 : The vast majority of L1 elements in the human genome are 5' truncated... 6

Figure 1.3 : The human and mouse model active L1 elements used for experiments... 8

Figure 2.1 : Three α-mouse ORF1 rabbit monoclonal antibodies characterized by immunoprecipitation (IP), immunoblot, and epitope mapping... 16

Figure 2.2 : Development of four novel mouse models that express tagged ORFeus-Mm L1 elements for interactome analysis................................ 21

Figure 2.3 : Selection of tagged mouse ORF1 and ORF2 constructs that retrotranspose at near-wild-type levels and the corresponding final constructs engineered for introduction into mice........................... 23

Figure 2.4 : Initial tests to solubilize and isolate GFP protein from cryo-milled grindate prepared from mouse tissue... 30

Figure 2.5 : An illustration of Thor pulverizing cryo-milled cells with his hammer to isolate L1 RNP complexes... 33

Figure 3.1 : L1 architecture and the design of the trialanine scan...................... 44

Figure 3.2 : The stammer in the coiled-coil of ORF1p introduces a flexibility that has inspired a theory for the formation of higher order structures by ORF1p homotrimers.. 46

Figure 3.3 : The customized L1 construct and the build of the 538 trialanine variants 48

Figure 3.4 : Detailed high-throughput cloning procedure used to create library 51

Figure 3.5 : Restriction digest used to validate final constructs in library... 53

Figure 3.6 : The retrotransposition efficiency of each trialanine variant............. 62

Figure 3.7 : High-throughput microscopy-based measurement of retro-transposition.. 64

Figure 3.8 : Retrotransposition levels mapped onto EN 3D structure................ 67

Figure 3.9 : Immunoblots of variants of ORF1p.. 70

Figure 3.10 : Impact of ORF1 mutations on protein abundance........................ 71

Figure 3.11 : Retrotransposition and protein abundance of mutants mapped onto WT ORF1p crystal structure... 72

Figure 3.12 : Schematic of the *cis*-preference model for binding of L1 RNA 79

Figure 3.13 : Protocol for measuring an ORF1p variant's ability to form RNPs using pooled screening.. 80

Figure 3.14 : RNA abundance ORF1p variants for RNP studies (*Experiment 1*)... 83

Figure 3.15 : RNP formation of each ORF1p variant for RNP studies (*Experiment 1*)... 84

Figure 3.16 : Functional categories of ORF1p mutants based on RNP studies (*Experiment 1*)... 85

Figure 3.17 : RNP formation of each ORF1p variant for RNP studies (*Experiment 2*)... 89

Figure 3.18 : IF analysis reveals intriguing nucleolar localization of some ORF1p variants.. 96

Figure 3.19 : Trends in amino acid conservation and sensitivity to mutation across ORF2p...108

List of Tables

Table 2.1 : 30 rabbit monoclonal anti-mouse ORF1 antibody subclone IDs………. 15

Table 2.2 : Tagged mouse ORF1 and ORF2 constructs tested for retrotransposition efficiency………………………………………………………….. 22

Table 2.3 : Overview and status of tagged mouse ORFeus transgenic mouse lines…………………………………………………………….…. 27

Table 3.1 : Organization of cloning trialanine variants into nine chunks…………. 49

Table 3.2 : Plasmids made in work with trialanine library…………………………. 50

Table 3.3 : Raw retrotransposition efficiency data for full library………………….. 56

Table 3.4 : Impact on retrotransposition efficiency organized by protein domain ………………………………………………………………………. 66

Table 3.5 : Categories of ORF1p variants based on activity and protein abundance…………………………….…………………………………….. 73

Table 3.6 : Composition of mutant pools for RNP studies (*Experiment 1*)……………………………………….……………………… 82

Table 3.7 : RNP formation of each ORF1p variant for RNP studies (*Experiment 2*) ……………………….…..…………………………………..88

Table 3.8 : An analysis of ORF1p variants that localize to the nucleolus…………. 94

Table 3.9 : IDs of fifty-five LINE-1 ORF2p sequences used for phylogenetic conservation analysis……………………………………………………….. 99

Table 3.10 : Phylogenetic conservation values of ORF2p variants used for sensitivity analysis ………………………………………………………100

Page intentionally left blank.

Chapter 1

Introduction to LINE-1 retrotransposition

The long interspersed element-1s (LINE-1s or L1s) are mobile genetic elements that use a "copy and paste" mechanism called retrotransposition to propagate themselves within the host genome. Ongoing L1 activity means that retrotransposition continues to shape the evolution of mammalian genomes [1–3]. See Figure 1.1 for a schematic of both the L1 construct used in our lab (top) and the L1 life cycle (bottom). Approximately 45% of the human genome is made up of retroelements, three of which are highly active families in modern humans: LINE1 (L1), *Alu* and SVA. About 17% of the human genome maps to L1 sequence [4], which includes roughly 500,000 copies of L1, the vast majority of the which are severely 5' truncated and are incapable of retrotransposition [5,6]. The truncation pattern strikingly holds true and is not well understood mechanistically (Figure 1.2). Approximately 90 L1 elements per diploid genome remain retrotransposition-competent, and L1 the only autonomously active mobile element. [7,8]. *Alu* and SVA elements depend on L1-encoded proteins to execute their retrotransposition in genomes and are thus considered non-autonomous.

By studying L1 retrotransposition in human tissue culture as well as in mice and rats, two widely used mammalian models for retrotransposition, there is evidence that L1 activity is highest is in the germline. Interestingly, somatic insertion events also occur in a variety of tissues, especially the brain, as well as during early development [9,10]. Other than the many examples of insertions into coding regions causing human disease [11], L1 is linked to tumor development in various cancers. ORF1p expression is elevated in many cancers, including breast, ovarian, and pancreatic cancers [12]. L1 insertions in genes such as *APC* have served as driver mutations in cancer, and the more metastatic cancer samples are studied, the more the field has appreciated that the environment of cancer cells supports L1 expression and retrotransposition activity [13–17]. Heightened L1 activity has also been reported to correlate with aging, stress, DNA damage, and telomere shortening, all of which likely work to keep the mutagenic capacity of L1 jumping in check [18].

The full-length human L1 element specifies production of a ~6kb long transcript that encodes two proteins, ORF1p and ORF2p [7], has a bidirectional promoter in its 5'UTR [19,20], and has a 3'UTR containing a weak polyadenylation signal [21–23]. ORF0, a 71-amino acid primate-specific ORF, has recently been described, and is transcribed antisense to ORF1 and ORF2 from within the L1 5'UTR; it may weakly promote retrotransposition [24]. The ORF1 and ORF2 proteins are both essential for retrotransposition and have some well-described biochemical activities. ORF1p is a 40 kDa protein that has both RNA-binding and RNA chaperone activities [25,26]. ORF2p is a 150 kDa protein that has endonuclease [27], reverse transcriptase [28], and nonspecific nucleic acid binding [29] activities. Upon translation of L1, ORF1p and ORF2p are thought to bind the RNA molecule from which they were transcribed through a poorly understood process called *cis*-preference thought to require the 3' poly(A) tail of L1 RNA [21,30,31]. ORF1p is translated quite efficiently, but ORF2p translation occurs at much lower levels, through an unconventional process that is also poorly understood [32]. The L1 RNA, ORF1p, ORF2p, complex is referred to as the L1 ribonucleoprotein (RNP) complex and is likely to be the direct intermediate of retrotransposition [33–38]. L1 insertion at the target genomic locus occurs via target-primed reverse transcription (TPRT) [39,40].

The L1 RNP composition is complex and dynamic in that its intracellular location and composition changes throughout the L1 its lifecycle [37,38,41]. There are mechanisms that inhibit such potentially mutagenic events as well as some that promote a complete L1 insertion event. A lot of research has gone towards identifying and characterizing these factors. This has mainly been done using both knockdown screens that measure the impact of a given gene on L1 activity in tissue culture as well as a few elegant proteomic approaches, in which L1 RNPs are isolated under various conditions and the interactors are identified using mass spectrometry [37,38,42–51].

For both mouse and human, the specific endogenous L1 sequence copies we use are taken from the organism's genome and are known to be retrotransposition competent. For both human (the copy used in our studies is named L1-rp, accession number AF148856) and mouse (the copy used in our studies is named L1-SPA, accession number AF016099), this includes the native 5'UTR, ORF1 and ORF2 protein coding sequences (and inter-ORF regions) as well as the 3'UTRs. Figure 1.3 compares the two elements. We also have a both mouse and human ORFeus constructs (ORFeus-Mm and -Hs, respectively) corresponding to each of them [52,53]. These are synthetic elements that have been recoded for maximum activity. We also tend to use non-endogenous 5'UTR/promoter sequences in most of our engineered systems.

There is still much more that remains to be understood about the how L1 completes its lifecycle. We have a great deal to learn from model organisms, which can provide both tissue-specific and developmental state information on the formation of RNPs. Also, since both ORF1p and ORF2p are essential for retrotransposition, a comprehensive picture of how the various domains and motifs of the full-length proteins contribute to the lifecycle is of great interest.

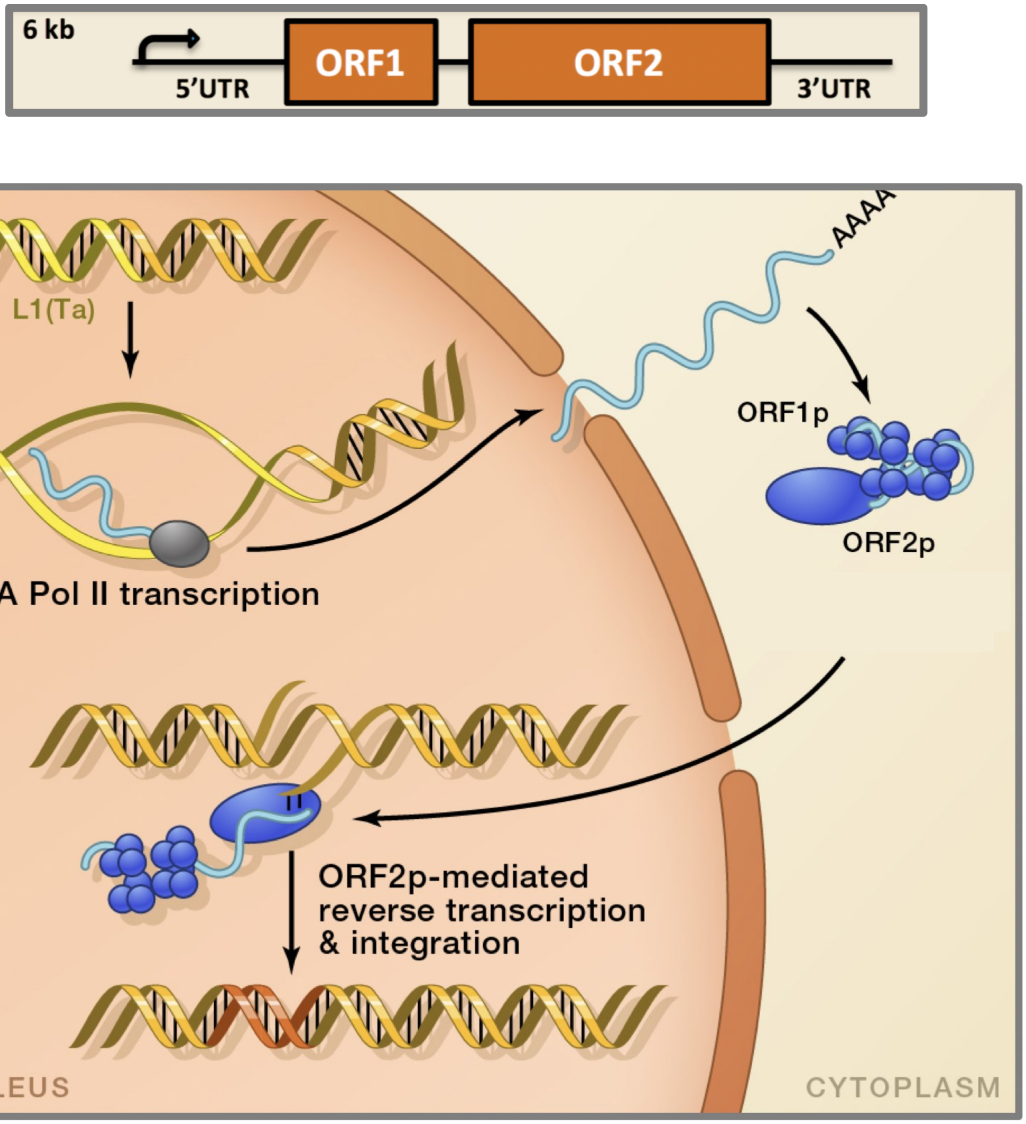


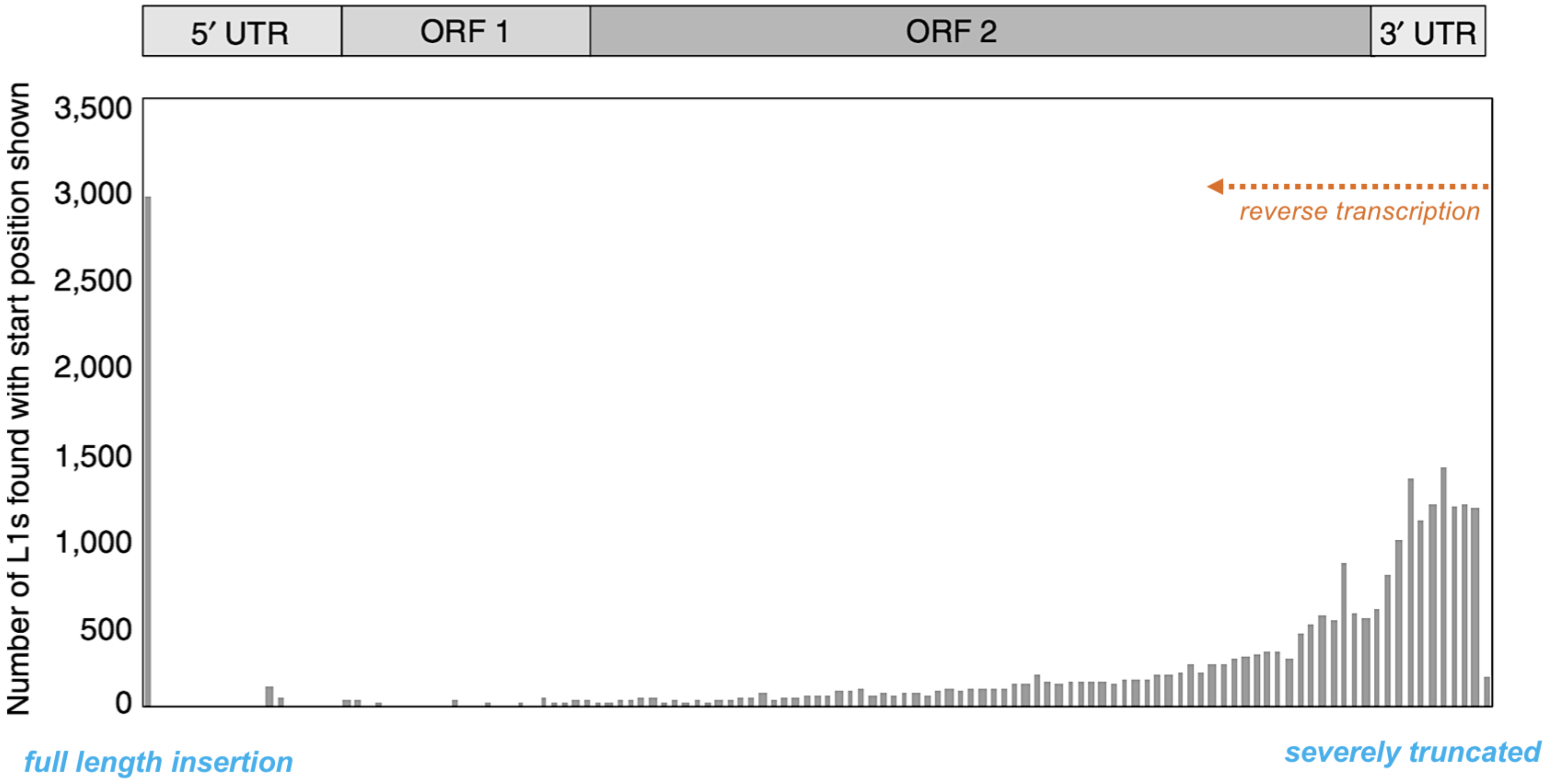
Figure 1.1 : The structure and lifecycle of L1.
The top shows a simplified structure of L1, which encodes the ORF1 and ORF2 open reading frames. There is a promoter in the 5’UTR and a polyadenylation sequence in the 3’UTR. The bottom shows a simplified schematic of the L1 lifecycle, which occurs via a copy and paste mechanism. The L1 mRNA, ORF1, and ORF2 proteins assemble (along with a variety of host factors) to form ribonucleoprotein complexes (RNPs) that cooperate in the L1 retrotransposition lifecycle.

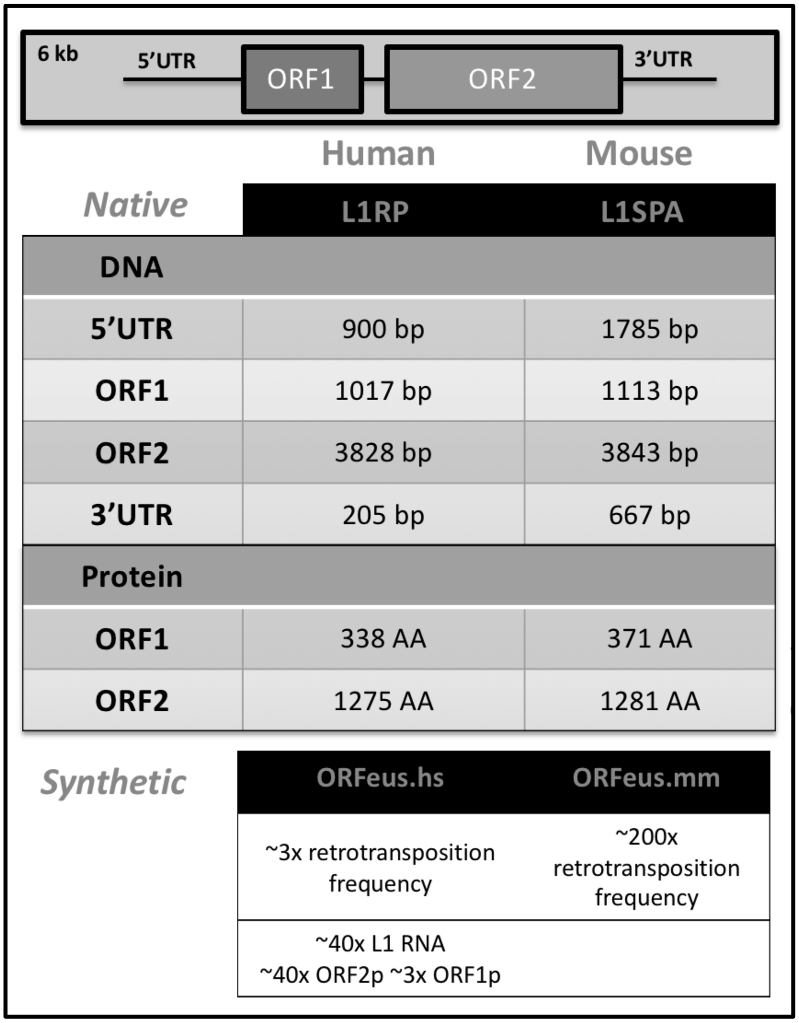
Image adapted from:
K. Burns and J.D. Boeke, Human Transposon Tectonics, Cell (2012).

Figure 1.2 : The vast majority of L1 elements in the human genome are 5' truncated.
Although 17% of the human genome (most recently appreciated to have ~500,000 copies of L1) maps to L1, over 90% of the L1 elements are not full-length and are severely 5' truncated. This chart highlights the frequency of the lengths of L1 elements. The schematic of an L1 element at the top indicates the position in the L1 element. As indicated in orange, it helps to visualize the reverse transcription starting at the 3' end of the element. This chart shows the frequency of a full insertion at the far left of the graph and represents, as moving to the right, the frequency of increasingly 5' truncated elements.

Image adapted from :
S. Szak et al., Molecular archeology of L1 insertions in the human genome,
Genome Biology (2002).

See next page.


5′ UTR
ORF 1
ORF 2
3′ UTR
Number of L1s found with start position shown
3,500
3,000
2,500
2,000
1,500
1,000
500
0
reverse transcription
full length insertion
severely truncated



Native	Human	Mouse
	L1RP	L1SPA
DNA		
5’UTR	900 bp	1785 bp
ORF1	1017 bp	1113 bp
ORF2	3828 bp	3843 bp
3’UTR	205 bp	667 bp
Protein		
ORF1	338 AA	371 AA
ORF2	1275 AA	1281 AA

Synthetic	ORFeus.hs	ORFeus.mm
	~3x retrotransposition frequency	~200x retrotransposition frequency
	~40x L1 RNA ~40x ORF2p ~3x ORF1p	

Figure 1.3 : The human and mouse model active L1 elements used for experiments.
The simplified structure of the L1 element is shown on top. Below, human and mouse L1 elements are compared. We use constructs expressing the human L1-rp and mouse L1-spa sequences for experiments. The lengths of each component of the L1 elements are displayed for both the DNA and protein sequences. The synthetic constructs are called ORFeus-Hs and -Mm respectively, they have been engineered for both the human and mouse elements and have silently recoded DNA in the protein coding regions to increase retrotransposition and protein expression, as shown. The fold-increase of the synthetic elements relative to their wild-type counterparts is indicated.

Page intentionally left blank.

Chapter 2

Tool development for studying L1 RNPs formed *in vivo*

Summary

This work merges two substantial areas of work in the L1 field. The first is studying mammalian L1 expression *in vivo*. Several previously established mouse and rat models have offered a lot of insight into tissue-specific and temporal L1 expression patterns in rodents. The second is the emerging sophistication with which we can immunoprecipitate (IP) L1 RNPs formed in tissue culture. We intended to expand the tools available to efficiently and deeply study the L1 RNP tissue- and developmental- specific L1 interactomes in mice. We present the development of tools that are customized for intricate biochemical analysis of active L1 complexes in mice, which include three rabbit monoclonal anti-mORF1p antibodies as well as engineered mice that highly express a ORFeus-Mm transgene with either the mouse ORF1 or ORF2 protein (mORF1p or mORF2p) harboring a protein epitope tag.

Introduction

L1 is a substantial component of the human and mouse genomes. Yet, why L1 is still active and the implications of this activity remain a mystery. Retrotransposons use many host cell proteins, tapping into the host's existing pathways to achieve the mechanistic steps involved in replicating their genome and inserting it into host genomic DNA. A myriad of host proteins also counter ongoing retrotransposition. Discovering the host factors that interact with L1 is potentially of great value, for it will expand our understanding of the molecular mechanisms that underlie possible role of L1s in human development, neural plasticity, aging, and cancer.

Our lab has undertaken a large effort to characterize these L1 host factors though immunoprecipitation of active human L1 RNPs in human tissue culture [37,38]. In this work, we overexpressed tagged L1 proteins. The tags were added to the C-termini of either L1-ORF1p or ORF2p such that they did not impair retrotransposition. The tags are recognized by extremely well-established systems for antibody-epitope recognition that work very well for immunoprecipitation (IP) in diverse conditions. This work generated a list of high-confidence protein interactors and helped greatly in dissecting the composition of L1 RNPs.

Isolating L1 RNPs from a living animal is an exciting and unexplored direction in which to take this work. In addition to the proteomic work referenced above, we started this project with valuable insight into expressing engineered L1 cassettes in rodents. Many transgenic L1 mouse lines have been generated and studied : (i.) a mouse model that expressed human L1-rp with a fluorescent reporter that exhibited human L1 expression in the testis and ovaries as measured by RT-PCR [54], (ii.) a similar system was further characterized to show human L1 activity in the brain, including during adult and embryonic neurogenesis and *de novo* retrotransposition events that integrated during embryogenesis [9], (iii.) a separate system in which a strong constitutive promoter driving ORFeus-Mm exhibited retrotransposition in both the germline and somatic tissue at levels ~20-fold higher than native expression [55], as well as (iv.) lines in which the Tet (doxycycline inducible) promoter driving ORFeus-Mm exhibited tightly regulatable high levels of retrotransposition in somatic tissue (O'Donnell et al. 2013), and also other mouse and rat models both human L1RE3 and L1-rp transgenes ([57] and unpublished work). L1 insertion events have also been studied in non-transgenic mice [58]. Having such an expansive range of expression of L1 transgenes in mice gave us great confidence that our system design, described below, would likely be successful and robust. Our tagged lines will add to the variety of mouse models available and will be the only ones optimized for targeted IP of L1 proteins.

To accomplish isolating active L1 RNPs, we needed to optimize the system so that the L1-encoded proteins could be efficiently immunoprecipitated from mouse tissue. Because no prior antibodies against mORF1p and mORF2p that could accomplish this task well, we decided to go about this project two ways. First, we attempted to develop an anti-mORF1p antibody that would recognize (untagged) mouse ORF1 extremely efficiently for IPs and other biochemical analyses. Second, we planned to exploit already established protein epitope tag – antibody pairs, entailing introduction of these tags into mORF1p and mORF2p in the context of an L1 expression cassette that would subsequently be used to establish novel transgenic mouse lines.

We hope to gain insight into the diverse components that both promote and help inhibit the lifecycle of L1 particles in a complex mammalian system. To do so, we strove to build an optimized set of tools for isolating L1 RNPS *in vivo*. Here, we present the development of and status of these novel reagents for use in studying retrotransposition in mice, an established and powerful animal model. As described, the tools described here were originally developed to accomplish the first attempt at looking at active L1 molecular complexes that form in a living organism, but they have already been and could be useful in myriad other applications.

Results and Discussion

- ***Production of anti-mouse ORF1 antibodies***

Full-length mORF1 was purified (see Methods) and sent to Abcam for the production of rabbit monoclonal antibodies. They conducted the injection of antigen into rabbits and sent the following samples back to us for testing at three different phases in the process : the polyclonal antiserum, multiclone supernatants, and finally the monoclonal subclone supernatants. Not all levels of testing are reported here. Most importantly, in the final steps of the process we tested 30 monoclonal subclones (Table 2.1) for both immunoblot and IP efficiency and chose three monoclones. We also mapped the epitopes recognized by these antibodies through a company called Pepscan. (Figure 2.1). Many milligrams of pure antibody were produced for each of these three antibodies for use in experiments. Between work in our lab and in those of collaborators, it has been demonstrated that these antibodies work very well for Western blot, immunohistochemistry, and IP analyses of ORF1p in mouse tissue (both publications have been accepted and are in currently in press : (1) Prostate-specific loss of UXT promotes cancer progression. Yu Wang, Eric Schafler, Phillip Thomas, Susan Ha, Gregory David, Emily M. Adney, Michael Garabedian, Peng Lee, Susan Logan, *Oncotarget, in press.* (2) LINE-1 elements are derepressed in senescent cells and elicit a chronic Type-I Interferon response. Marco De Cecco, Takahiro Ito, Amy E. Elias, Nicholas J. Skvir, Steven W. Criscione, Alberto Caligiana, Greta Brocculi, Emily M. Adney, Jef D. Boeke, Jayakrishna Ambati, Matthew Simon, Andrei Seluanov, Vera Gorbunova, Eline Slagboom, Stephen L. Helfand, Nicola Neretti, John M. Sedivy, *Nature, in press.)* Since the epitopes are known, we have also ordered peptides to be used for native elution in IPs: *mORF1pep01* for use with for use with abEA02 (Ac-NLDLDLKAYLM-

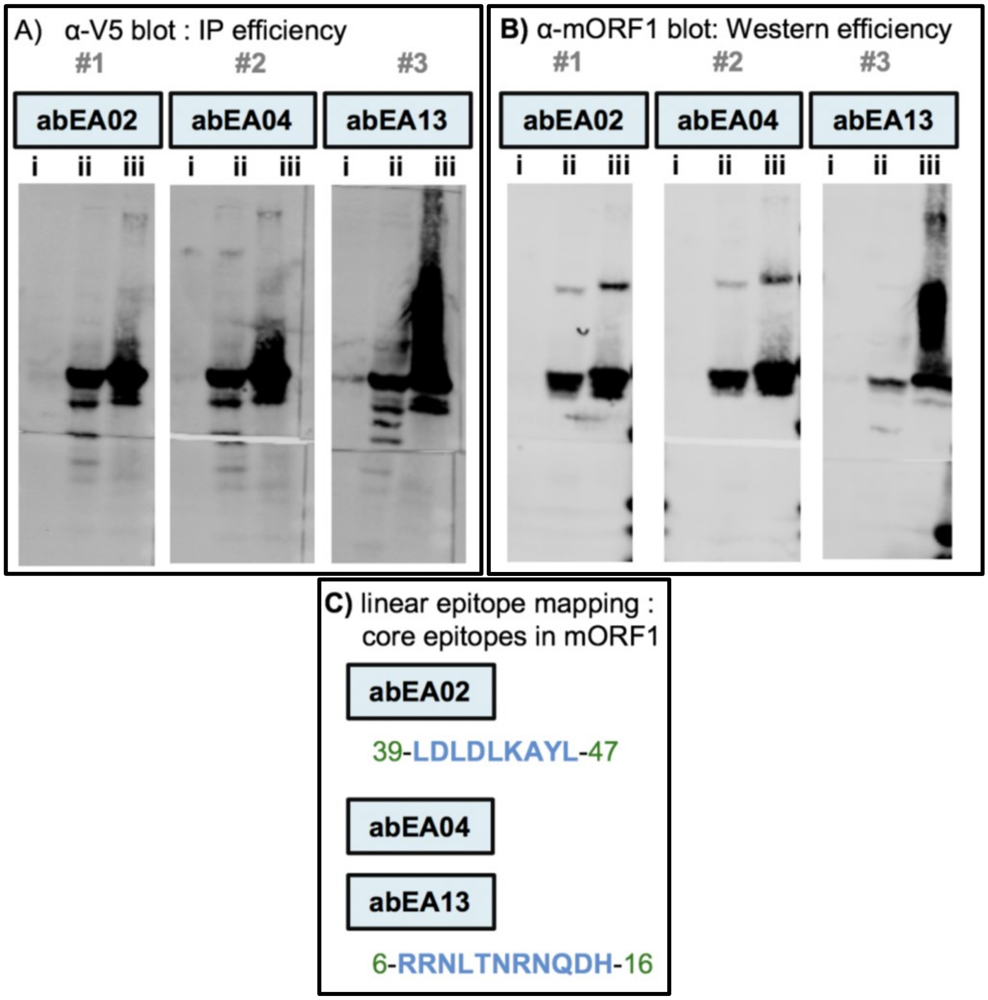
PEG_4-NLDLDLKAYLM-amide) and *mORF1pep02* for use with abEA04 and abEA13 (Ac-RRNLTNRNQDH -PEG_4-RRNLTNRNQDH-amide). However, the ability of these peptide antigens to elute ORF1 from antibody has not yet been evaluated. We would have used the same approach for anti- mORF2p antibody production, however we were unable to make enough antigen (mORF2p) to send for effective injection into rabbits.

Abcam ID	My ID #
1-1	1
1-2	2
1-4	4
1-5	5
1-6	6
1-7	7
1-9	9
1-10	10
1-11	11
1-12	12
16-1	13
16-2	14
16-3	15
16-5	17
16-6	18
16-7	19
16-8	20
16-9	21
16-10	22
16-11	23
16-12	24
23-1	25
23-2	26
23-3	27
23-4	28
23-5	29
23-7	31
23-8	32
23-11	35
23-12	36

Table 2.1: 30 rabbit monoclonal anti-mouse ORF1 antibody subclone IDs. Working with Abcam, we ended up with 30 candidate monoclonal antibodies. Detailed testing and data for such is provided in my Benchling and written notebooks. This table provides the IDs for these antibodies. Hybridoma cells for each exist in the liquid nitrogen freezer storage.

Figure 2.1 : Three α-mouse ORF1 rabbit monoclonal antibodies characterized by immunoprecipitation (IP), immunoblot, and epitope mapping.
We developed three rabbit monoclonal antibodies against mouse ORF1 (mORF1): abEA02, abEA04, and abEA13. **A and B :** Supernatants of the monoclonal hybridoma cell lines (containing high levels of α-mORF1 antibody) were used to test antibody IP and Western efficiency. Cell lysate was obtained from HEK293T cells with and without overexpression of V5-tagged mORF1. For each panel (#1-3) there are three protein samples loaded : (i) 30μg protein in cell lysate with no V5-mORF1 expressed, (ii) 30μg protein in cell lysate containing overexpressed V5-mORF1, and (iii) the elution after IP with the corresponding α-mORF1 antibody from 600 μg protein in cell lysate containing overexpressed V5-mORF1. **(A)** α-V5 blot : (ii) shows the amount of V5-mORF1 and (iii) shows the efficiency of IP by the α-mORF1 antibody. **(B)** α-mORF1 blot : (i) shows non-mORF1 background signal (ii) shows the amount of mORF1 recognized by the α-mORF1 antibody. Among these three α-mORF1 antibodies, we have great reagents for mORF1 detection by both Western and IP. **(C)** Linear epitope mapping (PepScan) for each of the 3 antibodies provided insight into where each antibody recognizes mORF1. The amino acid sequence of the core epitope (blue) as well as the corresponding amino acid range (green) within mORF1 are shown. All three MAbs recognize the N-terminus. The core epitopes of abEA04 and abEA13 are the same. The epitope mapping information helps enable immunoprecipitation of mORF1 followed by native elution, obviating the need for tagged ORF1.

See next page.


A) α-V5 blot : IP efficiency
#1
#2
#3
abEA02
abEA04
abEA13
i ii iii
i ii iii
i ii iii
B) α-mORF1 blot: Western efficiency
#1
#2
#3
abEA02
abEA04
abEA13
i ii iii
i ii iii
i ii iii
C) linear epitope mapping : core epitopes in mORF1
abEA02
39-LDLDLKAYL-47
abEA04
abEA13
6-RRNLTNRNQDH-16

- ***Establishment of tagged L1 transgenic mouse lines***

One goal of our research is to understand which molecular components of the mouse cells that comprise living tissues impact the activity of L1 transposons. Mice expressing “tagged” L1 elements, in which an exogenous epitope (with an already well-established antibody pair) is added to a protein with or without a linker sequence, would make it possible to do the ultimate desired biochemistry. Using these pre-established high-affinity reagents to IP tagged mORF1p and mORF2p could be further optimized to recover biologically active L1 complexes from mouse tissues.

Thus, we designed four transgenic mouse strains in which one of the two proteins (ORF1p or ORF2p) was tagged, and one of two promoters was used (CAG or TET, replacing the endogenous 5’UTR). Our plan was to have the optimal tagged-L1 elements (see below) expression under the control of two different, powerful expression systems. In these four lines, the L1 element would either be under a constitutive promoter (called “CAG”) or a minimal inducible-CMV promoter (called “TET”). For the CAG promoter system, the L1 transgene will be expressed in many genetic backgrounds. For the TET promoter system, in order for the L1 element to be expressed, the mouse would also need to express an rtTA transgene (it is not a natural mouse gene), as well as have the small molecule doxycycline in its system (which can be administered through the drinking water). The four lines to be made are summarized in Figure 2.2.

Tagging proteins can severely impact activity if not designed and tested well. Thus, the linker - tag sequences were chosen based on their ability to retain the retrotransposition efficiency of the untagged mouse L1 construct. We cloned many linker-tag sequences onto the C-terminus of each ORF1p and ORF2p in constructs that encoded a reporter for retrotransposition efficiency and measured them in human tissue culture. We tried combinations of three well-characterized epitope tags: 3xFlag, 2xV5, and mCherry. We also tested the presence of rigid and flexible linkers of two lengths. These epitope tags were chosen because they formed pairs with extremely high affinity antibodies, we have had successful experience with then in IPs from L1 elements in human tissue culture, and they are conducive to native elution (allowing elution under non-denaturing conditions). The mouse L1 cassettes that were constructed and tested are shown in Table 2.2**.** The corresponding retrotransposition efficiency levels for each cassette as well as those that were designed for the final establishment of mouse lines, via recombination at the *Rosa26 locus* (a well-established and often-used locus that supports high, stable expression across all mouse tissues)*,* are shown in Figure 2.3.

Because a key part of this project is to have control over the activity of the L1 transgene, this project entails optimization of how we combine the mice expressing Tet-L1 elements with the rtTA transgene and varying amounts of doxycycline. In order to have expression of the Tet-L1 transgenes, they must be crossed with mice expressing the rtTA gene. These lines must carry at least one copy of each transgene (L1 and rtTA) to start studies of L1RNPs. Since CAG is a strong constitutive promoter that we have experience with in other transgenic models of L1 activity that we have made, we expect L1 expression to be high. These lines will be valuable because we know some tissues may not express enough rtTA or be ideal for delivery of doxycycline with certain dox-rtTA transgene inducible systems.

As of now, we have two lines established suitable for IP of L1 proteins. Table 2.3 summarizes which lines we have established, along with some important notes, and the Methods expands further on the efforts to make these mice. The CAG-tagged mORF1p line and the Tet-tagged mORF2p (+ rtTA, double heterozygote) are now ready for IPs and proteomic analysis. We are on our way to understanding the interactome of L1 RNPs formed *in vivo*, the energy of which is captured in an illustration made to represent our RNP studies (Figure 2.5).

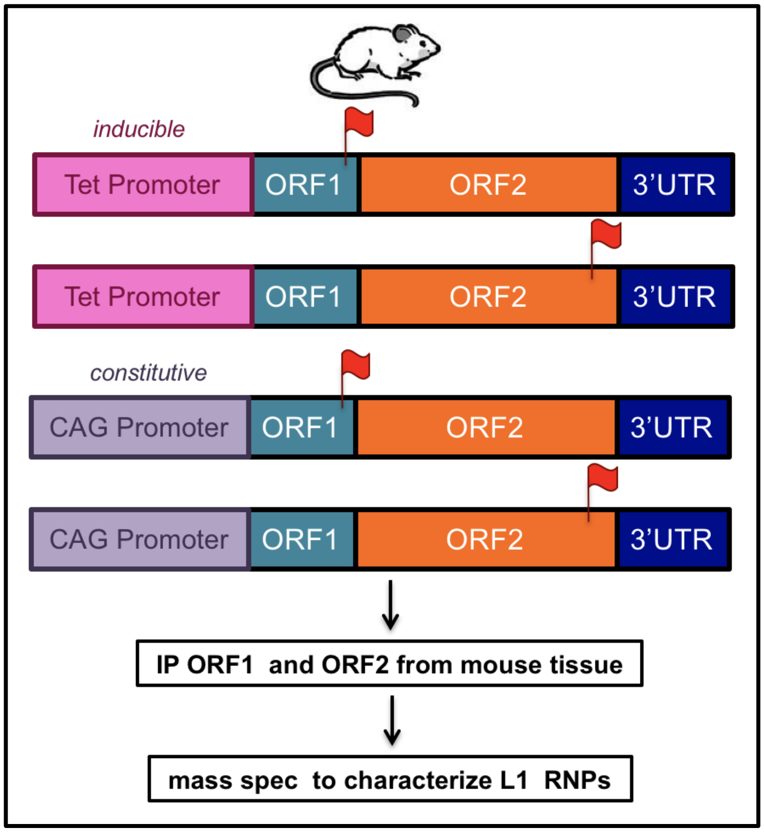


Figure 2.2 : Development of four novel mouse models that express tagged mouse ORFeus-Mm L1 elements for interactome analysis.
In an effort to isolate active L1 RNPs from live mouse tissue, we sought to engineer mice with optimal expression of mouse ORFeus that had ORF1p or ORF2p tagged under the control of either a constitutive of inducible (non-endogenous) promoter.

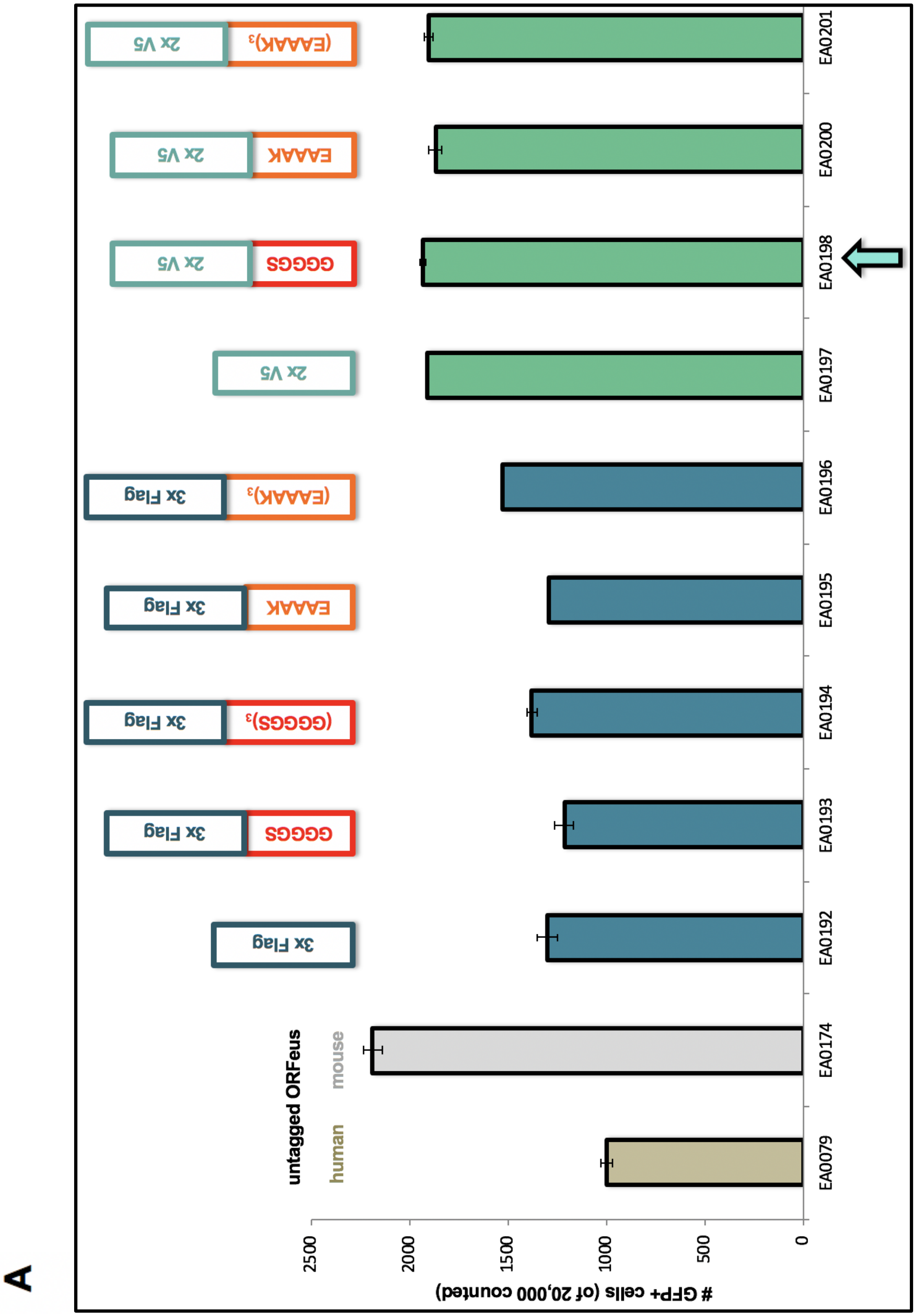
ORF1			
linker	tag	pEA0### ID	retroT % of untagged
none	3x-Flag	192	60
GGGGS		193	56
(GGGGS)$_3$		194	63
EAAAK		195	59
(EAAAK)$_3$		196	70
none	2xV5	197	87
GGGGS		198	88
EAAAK		200	86
(EAAAK)$_3$		201	87

ORF2			
linker	tag	pEA0### ID	retroT % of untagged
none	3x-Flag	213	96
GGGGS		214	87
(GGGGS)$_3$		215	100
none	2xV5	223	77
GGGGS		224	84
(GGGGS)$_3$		225	88
EAAAK		226	91
(EAAAK)$_3$		227	90
none	mCherry	228	89
GGGGS		229	87
EAAAK		231	85

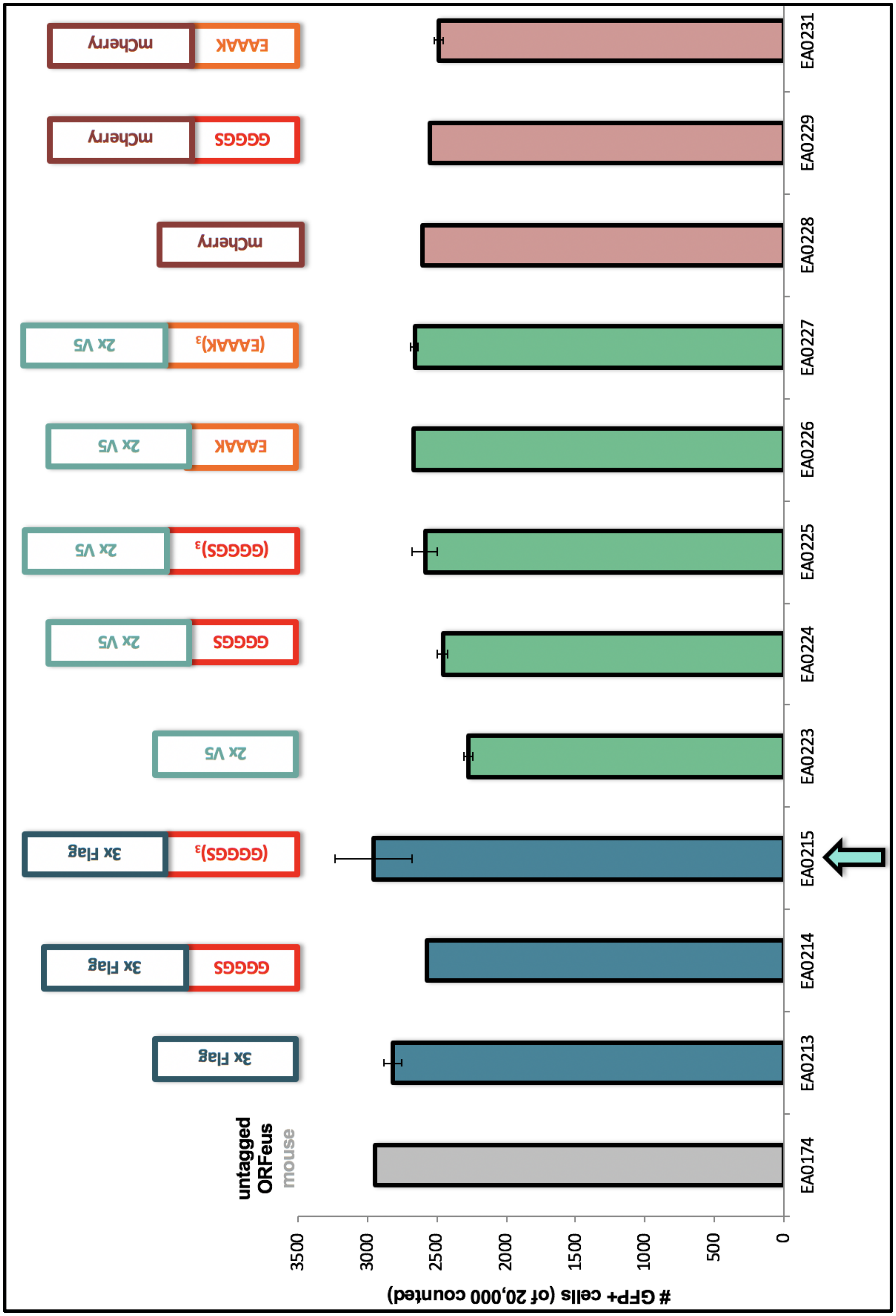
Table 2.2 : Tagged mouse ORF1 and ORF2 constructs tested for retrotransposition efficiency.
The tagged ORFeus-Mm constructs described above were cloned and expressed human tissue culture for measurement of retrotransposition. Only the C-termini were tagged, based on prior work. Tagging ORF1p (left) and ORF2p (right) is shown. For each protein : the left column indicates the linker amino acid sequence (no linker and both long and short flexible and rigid linkers were tested), the second column indicates the Tag, the third column indicates the pEA clone ID number, and the last column lists the average retrotransposition efficiency of the given construct, normalized to the activity of the untagged ORFeus-Mm.

Figure 2.3 : Selection of tagged mouse ORF1 and ORF2 constructs that retrotranspose at near-wild-type levels and the corresponding final constructs engineered for introduction into mice.
The constructs that showed the levels of retrotransposition closest to the wild-type were chosen and further cloned into plasmids designed for efficient insertion into the *Rosa26* locus in mouse blastocysts. (A and B) show the relative retrotransposition efficiency of the tagged ORF1p and ORF2p constructs, respectively. There are schematics of the C-term linker and tag above each construct. The bars of the graph represent the number of GFP+ cells, which represents retrotransposition activity. Indicated with a green arrow at the bottom are the constructs for which we chose the tagged L1 constructs for use in mice (pEA0198 for ORF1p and pEA0215 for ORF2p), which jumped at 88% and 100% the level of wildtype, respectively. (C) Schematics of the relevant L1 cassettes used to establish the mouse lines are shown (with the pEA clone ID numbers listed along the left). Aleks Wudzinska and Sangyon Kim contributed to the identification, breeding and characterization of the knock-in animals

Fig. 2.3 spans the next three pages.


A
# GFP+ cells (of 20,000 counted)
2500
2000
1500
1000
500
0
untagged ORFeus
human
mouse
EA0079
EA0174
3x Flag
EA0192
GGGGS
3x Flag
EA0193
(GGGGS)₃
3x Flag
EA0194
EAAAK
3x Flag
EA0195
(EAAAK)₃
3x Flag
EA0196
2x V5
EA0197
GGGGS
2x V5
EA0198
EAAAK
2x V5
EA0200
(EAAAK)₃
2x V5
EA0201

B



# GFP+ cells (of 20,000 counted)
0
500
1000
1500
2000
2500
3000
3500
untagged
ORFeus
mouse
EA0174
3x Flag
EA0213
GGGGS
3x Flag
EA0214
(GGGGS)$_3$
3x Flag
EA0215
2x V5
EA0223
GGGGS
2x V5
EA0224
(GGGGS)$_3$
2x V5
EA0225
EAAAK
2x V5
EA0226
(EAAAK)$_3$
2x V5
EA0227
mCherry
EA0228
GGGGS
mCherry
EA0229
EAAAK
mCherry
EA0231

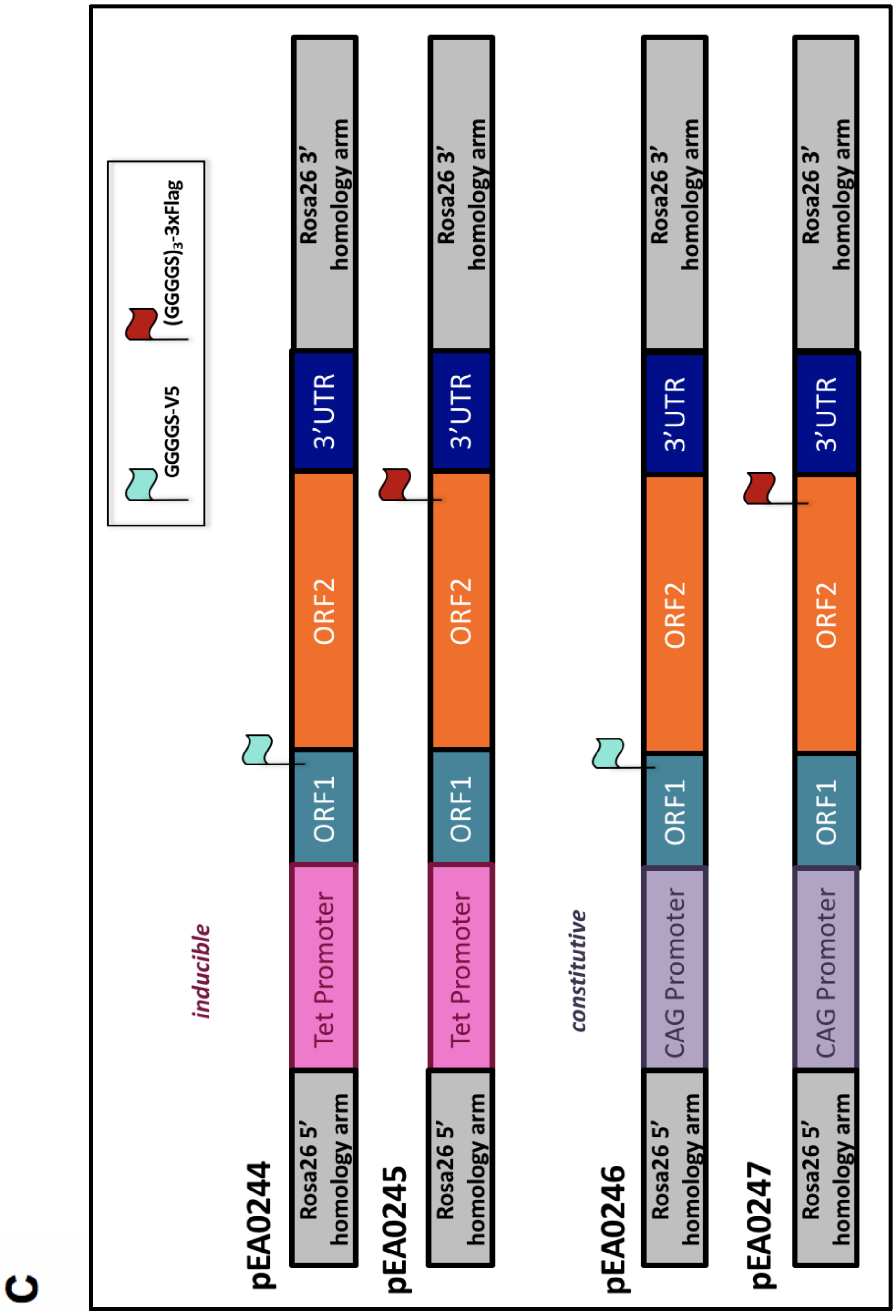

C
GGGGS-V5
(GGGGS)$_3$-3xFlag
inducible
pEA0244
Rosa26 5' homology arm
Tet Promoter
ORF1
ORF2
3'UTR
Rosa26 3' homology arm
pEA0245
Rosa26 5' homology arm
Tet Promoter
ORF1
ORF2
3'UTR
Rosa26 3' homology arm
constitutive
pEA0246
Rosa26 5' homology arm
CAG Promoter
ORF1
ORF2
3'UTR
Rosa26 3' homology arm
pEA0247
Rosa26 5' homology arm
CAG Promoter
ORF1
ORF2
3'UTR
Rosa26 3' homology arm

Table 2.3 : Overview and status of tagged mouse ORFeus transgenic mouse lines.
While we attempted to establish all four of the intended tagged mouse lines, we were successful in establishing two, as shown in this table. The first two columns indicate the L1 and rtTA genotype. The second column shows the dates upon which founder mice were obtained. The last column contains important notes on the lines that we are now working with to study the interactome in mouse tissues. All have the C57BL/6J background.

Mouse line genotype		**Founder mice date of birth (numberof mice)**	**Notes**		
L1	**rtTA**				***status***
CAGG-ORF1-V5 / -	- / -	12/25/6 (6 chimeras)	no L1 homozygotes born yet	backcrosses complete 5/18/2017	colony ready for proteomic analyses
Tet-ORF2-Flag / -	- / -	02/20/18 (6 chimeras)	no L1 homozygotes born yet	backcrosses complete 5/25/2018	N/A
Tet-ORF2-Flag / -	rtTA / -	08/10/2018 (2)	N/A	N/A	colony ready for proteomic analyses

- ***Initial tests of targeted isolation of protein from cryo-milled mouse tissue***

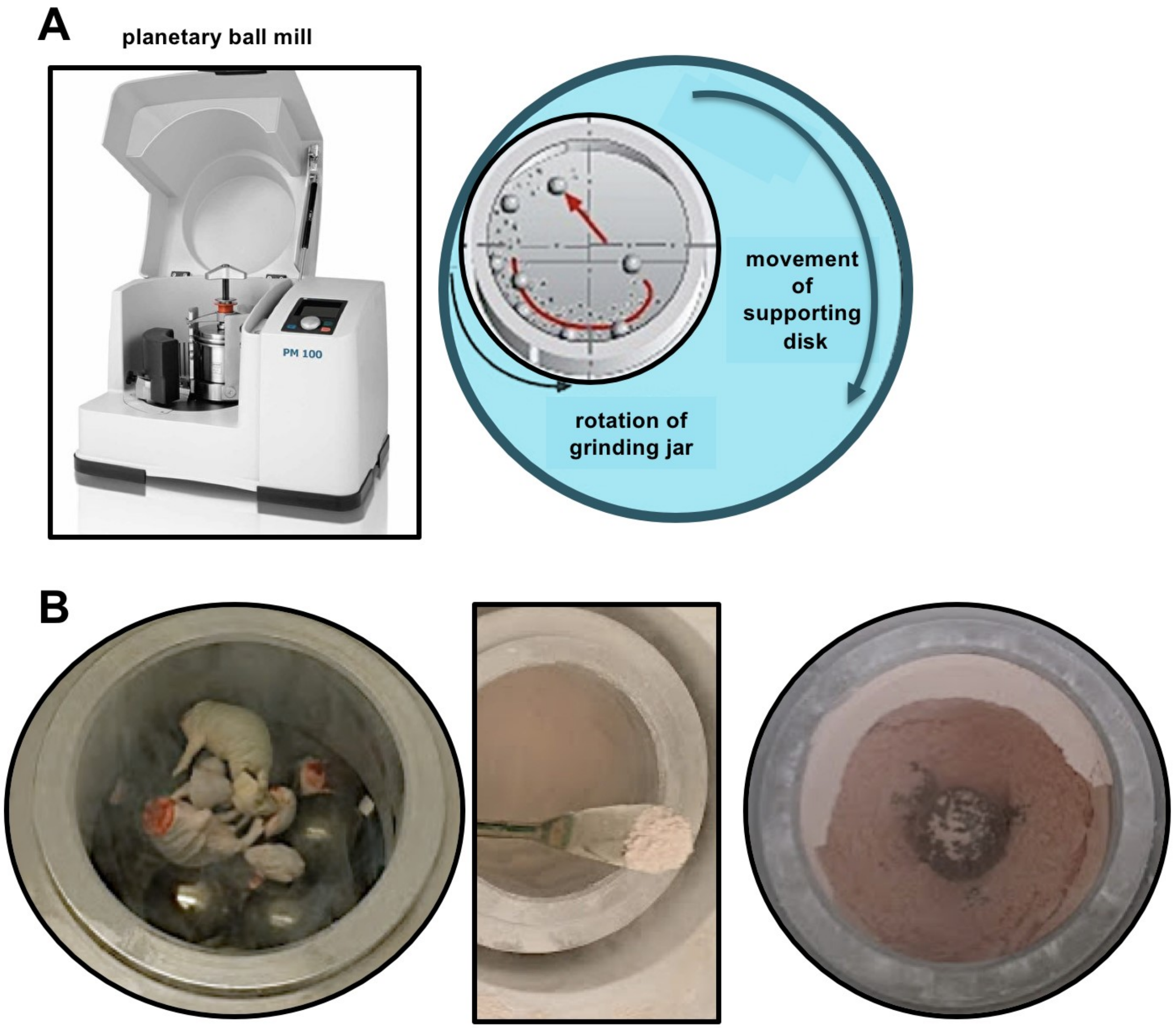
Isolation L1 RNPs formed *in vivo* requires preparation of the mouse tissue followed by IP of the targeted L1 proteins. In order to accomplish this, we needed to extend our expertise in doing this with human cells [37,38] to this much more complex, living mammalian system. For human cells in culture, the approach included flash freezing the cells in liquid nitrogen, lysing the cells in the form of this frozen material through cryo-milling to produce what is called grindate [59], solubilizing the material with optimized buffers (favoring the solubilization of the target protein complexes), conducting an IP targeting a protein of interest, and running these captures complexes through mass spectrometry analysis. We would ultimately like to take the same approach to doing this with mouse tissue, which necessitated a model system for starting to optimize each step.

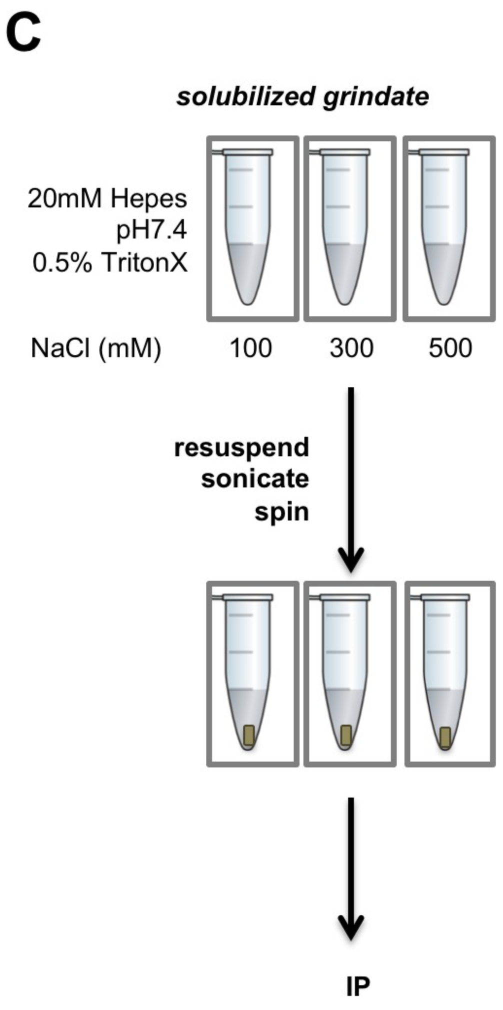
Although each protein (and any associated bound proteins) would likely require fine-tuned optimization at the extraction and IP steps, we started with initial tests to create initial workable protocol to IP GFP from tissue from mice ubiquitously expressing a GFP transgene (because we already had efficient reagents for IPs of GFP from human cells; see Methods and Figure 2.4). We demonstrated that we can create a grindate from mouse tissue and that under a series of buffers, could extract of soluble proteins, at high yield and across all sizes. We also successfully IP'd GFP from the soluble fraction (clarified lysate) prepared from mouse grindate, although only 16% of the total GFP in the clarified lysate was recovered in the IP. This was a nice proof-of-principle set of experiments that will help inform how to produce and productively work with grindate now that the appropriate tagged-L1 lines are established.

Figure 2.4 : Initial tests to solubilize and isolate GFP protein from cryo-milled grindate prepared from mouse tissue.

We maintained a mouse line that ubiquitously expressed a GFP transgene to use as a model for testing the extraction of a targeted protein (GFP) from mouse tissue grindate. (A) The left side displays the planetary ball mill instrument used, the closed vessel of which contained a tissue sample that was to be ground, the appropriate sized milling balls, and liquid nitrogen. The schematic on the right represents how the rotational forces are able to pulverize material in the vessel. (B) The pictures of the tissue at different points in the process to make the grindate : flash frozen, decapitated mouse pups (age day 3) (left) were cryo-milled to produce the grindate frozen powder (middle and right). (See Methods for details.) (C) Our first protein extraction experiment is outlined. Grindate was made using a short (S) or long (L) milling time. The two different grindate samples were solubilized in one of three buffers that differed only in the concentration of NaCl (in mM), indicated as either 100, 300 or 500 throughout the figure. There were 6 samples total prepared for IP. We span down the samples to separate the soluble (clarified lysate) from the insoluble (pellet) fractions and did an IP of GFP from the clarified lysate. We kept samples of the clarified lysate, the resuspended pellet, and the IP elution for analysis by gel. (D) A protein gel stained with Coomassie (top) shows the relative total protein obtained at each step for each condition described. The vast majority partitioned into the soluble fraction and increasing salt helped with the extraction, as expected. 500ng of bovine serum albumin (BSA) is shown as for a sense of protein concentration and where GFP ran on the gel is indicated with (< GFP). A Western analyses of GFP through the IP (bottom) shows (left) the relative absolute signal for GFP in the elution for each condition. The comparison of input to flow through for the ideal extraction condition (bottom, right) shows the GFP signal in the clarified lysate (Ly), pellet (Pe), and fraction unbound by IP (the flow-through, FT). This shows that the IP of GFP from mouse tissue grindate was successful, however only 16% of the total GFP in the clarified lysate was recovered by IP.

Figure 2.4 spans the next 2 pages.


A
planetary ball mill
PM 100
movement of supporting disk
rotation of grinding jar
B


C
solubilized grindate
20mM Hepes
pH7.4
0.5% TritonX
NaCl (mM)
100
300
500
resuspend
sonicate
spin
IP

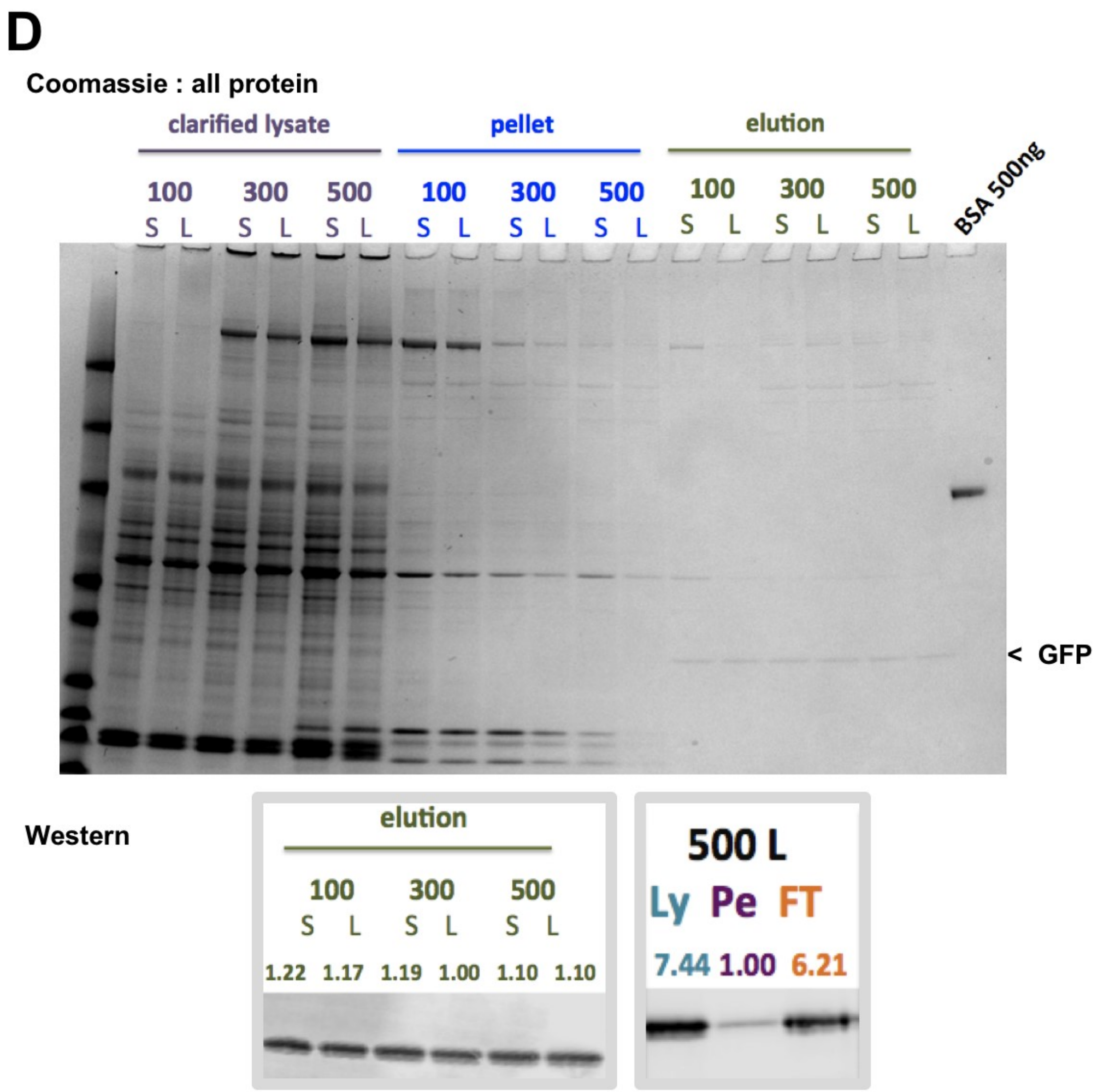

D
Coomassie : all protein
clarified lysate
pellet
elution
100 300 500
S L S L S L
BSA 500ng
< GFP
Western
elution
100 300 500
S L S L S L
1.22 1.17 1.19 1.00 1.10 1.10
500 L
Ly Pe FT
7.44 1.00 6.21

Figure 2.5 : An illustration of Thor pulverizing cryo-milled cells with his hammer to isolate L1 RNP complexes.
This was created by Sigrid Knemeyer. The team behind the work (Taylor 2013) collaborated to convey the imagery behind the work, mainly to capture the power of the cryo-milling and RNP isolation we had begun to study. This represents the concepts behind our work *in vivo* as well. (In fact, we actually needed to use a hammer to break up frozen mouse pups at one point, see Methods. We hope the hammer is not necessary in future, more developed protocols.)

Methods

Measuring the retrotransposition efficiency of the tagged ORFeus.Mm constructs

On day one of the experiment, 250,000 Tet-On HEK293T$_{LD}$ cells[46] were seeded per well in 2 mL DMEM (with 10% PBS and Penicillin-Streptomycin as described in Chapter 3) in a 6-well plate. 24 hours later, the DNA was transfected. For one well, a transfection mixture was made : 3 µL of Fugene-HD (Promega, cat #E3211), 1 µg miniprepped DNA, 100 µL Opti-mem. Once mixed well, we incubated the solution at room temperature for 25 minutes. Each well got the entire transfection mixture, evenly distributed, drop-wise. 24 hours later, the cells were split. We aspirated off the media, the cells were washed with 2 mL of PBS, and then 400 µL trypsin solution (TrypLE™ Express Enzyme (1X), Life Technologies cat # 12604039) was added the wells, and they were incubated for 5 minutes in the 37°C incubator. The plates were jarred by hand to dislodge cells, 600 µL of DMEM was added, and cells were resuspended by being pipetted up and down. 250 µL (25%) of the cells were transferred to 6 cm plates in a total of 4 mL (final puro 1 µg/mL) and 24 hours later, doxycycline was added in 100 µL DMEM (final dox 1 µg/mL). These plates then incubated for 3 days. The cells were then transferred to tubes for fluorescence detection by removing the media, adding 2 mL of PBS and breaking up cells by pipetting up and down. L1 retrotransposition frequency of constructs was measured by evaluating the signal of the GFP-AI reporter was evaluated by flow cytometry using a Becton Dickinson LSR II.

Production characterization and of anti-mouse ORF1 rabbit monoclonal antibodies :

Working with Abcam to produce RabMabs:

Full-length mouse-ORF1 protein was purified from bacteria in the lab of Kathy Burns by Marty Taylor and David Husband. In brief, the protein had a histidine-tag and a TEV cleavage site. It was purified using a nickel column, then treated a with an excess of TEV protease and RNAse to remove the tag, and then ran over a gel filtration column. The final concentration was 4 mg/mL in 0.5M NaCl, 20mM Phosphate pH 8, 10 mM $MgCl_2$, which purified as a monomer and was stable at lower [NaCl]. The protein was shipped to Abcam for injection into rabbits: For each injection, the maximum volume of the protein solution was 1 mL, which they then mixed this with 1 mL of adjuvant. The first injection was 0.4 - 0.5 mg of that 0.4 - 0.5 mg/mL. Later injections were 0.2 – 0.25 mg, so those can be of 0.2 – 0.25 mg/mL. They diluted the protein as needed in PBS, so we proved before we sent off the protein that ORF1p was stable in that buffer (data not shown). To make the description of the testing process on our as clear as possible, there is a detailed figure legend provided in Figure 2.1.

Working with Pepscan to map the epitopes:

We sent 50 μg of each of three final purified anti-ORF1p antibodies to Pepscan for epitope mapping. Pepscan used their default approach to linear mapping, which included synthesizing discontinuous, linear 15-mer peptides that were tested as peptide epitope mimics. These were generated by simply providing them with the mORF1 primary sequence. These peptides were made *in situ* and assembled on a mini-array. The binding of each of three monoclonal antibodies to the peptides was measured using ELISA assays. For the three monoclonal antibodies, they measured the binding of each antibody to all of the peptides in the array. The epitopes (the peptides that corresponded to the highest binding affinities) were then reported to us. It took 10 weeks.

Making the cryo-milled grindate from decapitated mouse pups :

We maintained a line of mice that constitutively expressed eGFP in essentially all tissues (JAX stock #: C57BL/6-Tg(CAG-EGFP)1Osb/; through crosses with C57BL/6 mice) When pups were born, we laid them on ice for 30 minutes and decapitated them according to our IACUC protocol. The heads and bodies were dropped directly into liquid nitrogen and allowed to flash freeze for one minute. The remaining liquid nitrogen was discarded and the mouse samples were stored at -80°C. Due to the genotype, we expected 75% of our mice to express eGFP. We used 6 (age day 3) pups total.

The samples were converted into cryo-milled grindate using a large planetary ball mill (Planetary Ball Mill PM 100) and tungsten balls. We followed an extremely well described protocol, also used for making our grindate frozen droplets of mammalian cells, which is extremely well described. [59] The only variation from this procedure was that the mouse tissue samples were quite large and needed to broken-down in to pieces about one tenth the size shown in Figure 2.4B, far left, which we did by putting the large pieces into a metal bowl with some liquid nitrogen. We broke them down manually, using a hammer and small metal chisel. We then ground this tissue in the ball mill using either a long or short interval (described below). In the future it would be advisable to cut up the pieces of mouse tissue before they are flash frozen so they can be ground efficiently. Having smaller pieces will also allow for more thorough and ultra-fast freezing, which will avoid the formation of ice crystals that might affect protein folding and proper complex isolation.

For the short grind, we used 125 mL jar and 5 x 20mm balls with liquid nitrogen and mouse tissue in the chamber. We did three sequential 3-minute grinds at 1 minute intervals, at 400 RPM. We then removed half of the material and subjected it to the longer grind, in which we used a 125 mL jar and 60 x 10 mm balls. We did two 1-minute grinds at 1 minute intervals, at 400 RPM. Both grindate samples were stored at -80°C.

We then added 400 µL of buffer (20 mM Hepes, pH7.4, 0.5% TritonX-100, Roche complete EDTA free tablet, added fresh, and either 100 mM, 300mM or 500 mM NaCl (see Figure 2.4C) to 100mg of grindate on ice. We resuspended by pipetting up and down and sonicated the samples[59], then spun all samples at 20,000 x g, 10 minutes, at 4°C. There was a large pellet and a substantial amount of lipid floating on top. We avoided that and took the middle soluble fraction for IP and gel analysis Figure 2.4D. For the IP we used a polyclonal llama anti-GFP antibody [60] that were coupled to Dynabeads M-270 epoxy concentration of 10 µg antibody/mg of Dynabeads (Invitrogen, cat. #143-02D). For the Western we used a different anti-GFP antibody (Roche Applied Science, cat. #11814460001).

Page intentionally left blank.

Chapter 3

Comprehensive scanning mutagenesis of the human retrotransposon L1 identifies novel motifs essential for function

Summary

To determine amino acid sequences in ORF1p and ORF2p that are critical for L1 function, we undertook a scanning mutagenesis study. We successfully assembled an ordered library of 538 trialanine variants that scan across both ORF1 and ORF2 proteins of a human L1 retrotransposon. We found that retrotransposition efficiency is extremely sensitive to ORF1p and ORF2p mutations, consistent with what one may expect from this streamlined and highly conserved element. The vast majority of ORF1p mutants form stable protein, but do not jump well. To characterize RNA binding affinity, we developed a novel sequencing approach to studying RNP formation. Using this approach, we show that most ORF1p variants efficiently bind L1 RNA. Some ORF1p variants showed a distinct nucleolar phenotype, the strongest of which overlapped with the stammer motif found in the coiled coil domain of ORF1p. Mutating the stammer motif also seemed to impair RNP formation. Mutating the RRM and CTD domains of ORF1 significantly impaired protein stability and likely RNP formation. We identify what we term the "Star Cluster" (as well as other regions in ORF2p), which is a dense series of residues that lie outside the well-studied regions of ORF2p and yet surprisingly appear to be important regions for function, that would not

have been predicted by conservation. These regions should be a high priority in the continued efforts to understand the non-catalytic domains of this protein.

Introduction

The autonomously active L1 is assumed to go through multiple stages to progress through its complex lifecycle and paste itself relatively efficiently into the human genome. This entails successful transcription of L1 RNA, protection of RNA from degradation, translation of ORF1p and ORF2p, folding and stability of these proteins, binding of the L1 RNA by these proteins to the L1 RNA (*in cis*), and ORF2-mediated TPRT at the targeted locus. This requires proper L1 cellular localization and ribonucleoprotein particle (RNP) – formation, which includes the assembly of L1 RNA, ORF1 and ORF2 proteins, as well as interactions with host proteins and RNAs throughout the L1 life cycle. Mutating L1 affects DNA, RNA, and protein primary sequences and, thus, may affect any of the steps listed above. By building and characterizing the first comprehensive scanning mutagenic library of any transposable element, we were able to begin comprehensive analyses of how disruption of L1 sequence may impact many of these cellular activities.

Elegant work that precedes this has provided structural data, targeted protein mutagenesis for functional studies, and RNA-binding analyses that have provided valuable insights into the wild-type functions as well as the specific sensitivities of individual ORF1p and ORF2p domains to amino acid alterations [27,28,61–64]. Despite success in discerning how regions of L1 proteins promote retrotransposition, as of yet, there is no comprehensive and unbiased analysis across both full-length protein coding regions.

We describe an ordered and comprehensive trialanine scanning mutagenic library of ORF1p and ORF2p in an active human L1 mobile element. To appreciate this study, it is essential to appreciate the topology of human ORF1p and ORF2p (Figure 3.1, top). ORF1p consists of an unstructured N-terminal region (NTR), followed by three structured domains, which include the coiled coil (CC) domain consisting of an extended series of heptad repeats, the RNA recognition motif (RRM) domain, and the C-terminal domain (CTD). The structure of human ORF1 has been well-characterized by x-ray crystallography [61,62], culminating in a near-full-length structural model used extensively in this report [62]. The CC domain causes ORF1p to trimerize [26,61], and the RRM and CTD domains are jointly responsible for single-stranded RNA-binding [61,65,66]. Recent work has shown that the extended CC domain structure is metastable as the consequence of a single “stammer” insertion (residues M91, E92 and L93) in the heptad repeat. This provides evidence for ORF1p homotrimers being able to sample both structured and partially unstructured states and this is thought to underlie its function [62]. This work from the Weichenrieder lab has expanded this flexibility to theorizing how ORF1p homotrimers may thus assume open and closed states, the open state may allow for functionally critical higher order structures. This proposed model is shown in Figure 3.2.

ORF2p also has regions of fairly well characterized structure and function. The most thoroughly understood regions functionally are the enzymatic endonuclease (EN) and reverse transcriptase (RT) domains [27,28]. Other less functionally defined, yet annotated motifs include the recently described Cryptic (Cry) sequence [63] and the Z domain region [67], and the carboxy-terminal segment (CTS), which harbors a cysteine rich motif [68] , which is important for retrotransposition. There is a crystal structure of the EN domain [64] but the majority of ORF2p remains structurally uncharacterized. In this work, I refer to two poorly characterized large regions of ORF2p as Desert 1 (the region between the EN and Z domains, which contains the Cry sequence) and Desert 2 (the region that lies after RT and contains the CTS and cysteine rich motif) (Figure 3.1, top).

This report focuses on how we built and characterized the library. For each of the 538 variants, we individually measured retrotransposition efficiency. In an effort to probe the genetic and physical interactions of L1 proteins, we studied the ORF1p variant portion of the library more deeply by measuring protein and RNA expression and started developing a method to probe each mutant's ability to form RNPs. For ORF2p, we provide a comparison of conservation and retrotransposition efficiency, helping identify which previously poorly characterized areas of ORF2p are of highest interest to study further. By combining these data, we parsed regions of ORF1p and ORF2p, by three-residue trialanine mutagenic segments, and organized them into functional categories. This library has been used to produce a comprehensive map that indicates which residues are critical or dispensable for the L1 lifecycle.

Results and Discussion

- ***Designing and building a trialanine scanning mutagenic library***

The L1-encoded proteins, ORF1p and ORF2p, consist of 338 and 1275 AA residues, respectively (Figure 3.1). To obtain a complete mutagenic scan of the coding sequences, we designed an ordered library 113 variants for ORF1p and 425 variants for ORF2, totaling 538 variants, each of which had three consecutive residues mutated to alanine (each referred to as a trialanine mutant). The mutants tiled along the proteins, did not overlap, and did not include the start or stop codons (Figure 3.1). Some exceptions occurred at the ORF termini due to “remainders” (see and Table 3.1 and *Methods*).

For an ordered library of this size, each of the 538 mutants needed to be built independently, necessitating a customized plasmid (pEA0264), which contained a wildtype (WT) L1 cassette with the addition of unique silent restriction sites, tailored to efficiently accommodate the introduction of each trialanine variant DNA fragment (Figure 3.3, Table 3.1) This design allowed for a simple and efficient two-piece assembly for each mutant construct. Each variant-containing DNA cassette was designed as a ~600bp fragment, contained the 9bp mutation with overlapping ends for assembly, and was synthesized (and provided by the company within a plasmid) (Figure 3.3). Assemblies were done in a 96-well format, with 95% efficiency of obtaining the correct clone when only one colony was checked (Figure 3.3-5). The plasmids used to make the library as well as the final constructs that made up the ordered library are detailed in Table 3.2 and the first column of Table 3.3.

Figure 3.1 : L1 architecture and the design of the trialanine scan.
The human L1 proteins are depicted in detail. The residue positions of characterized domains are shown for ORF1p and ORF2p. The library consists of 538 mutants. The design of the trialanine mutants for the first two and the last mutant of the library are shown at the DNA and protein sequence levels. Start and stop codons were not mutated. The trialanine variants are consecutive and non-overlapping.

See next page.

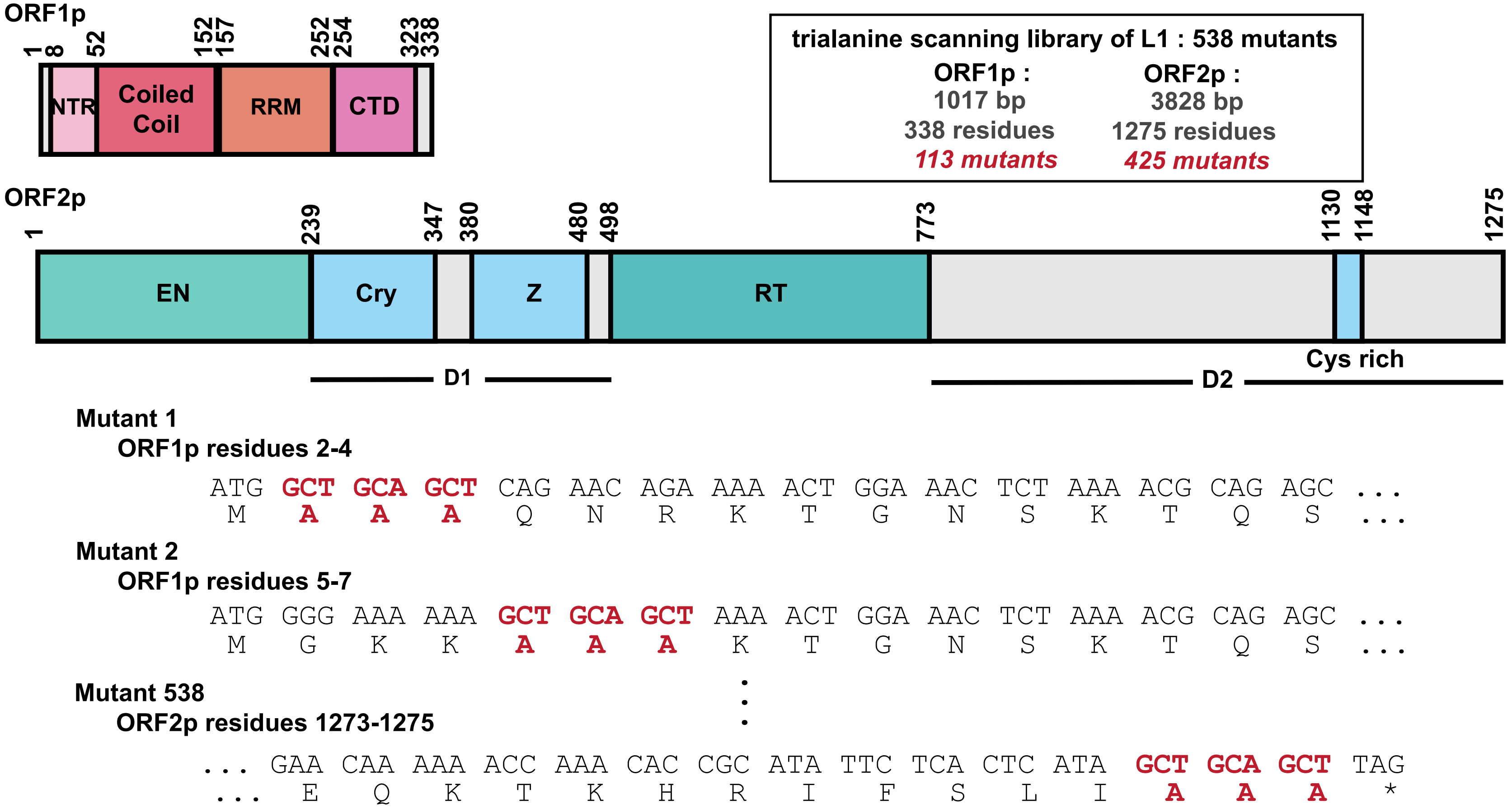
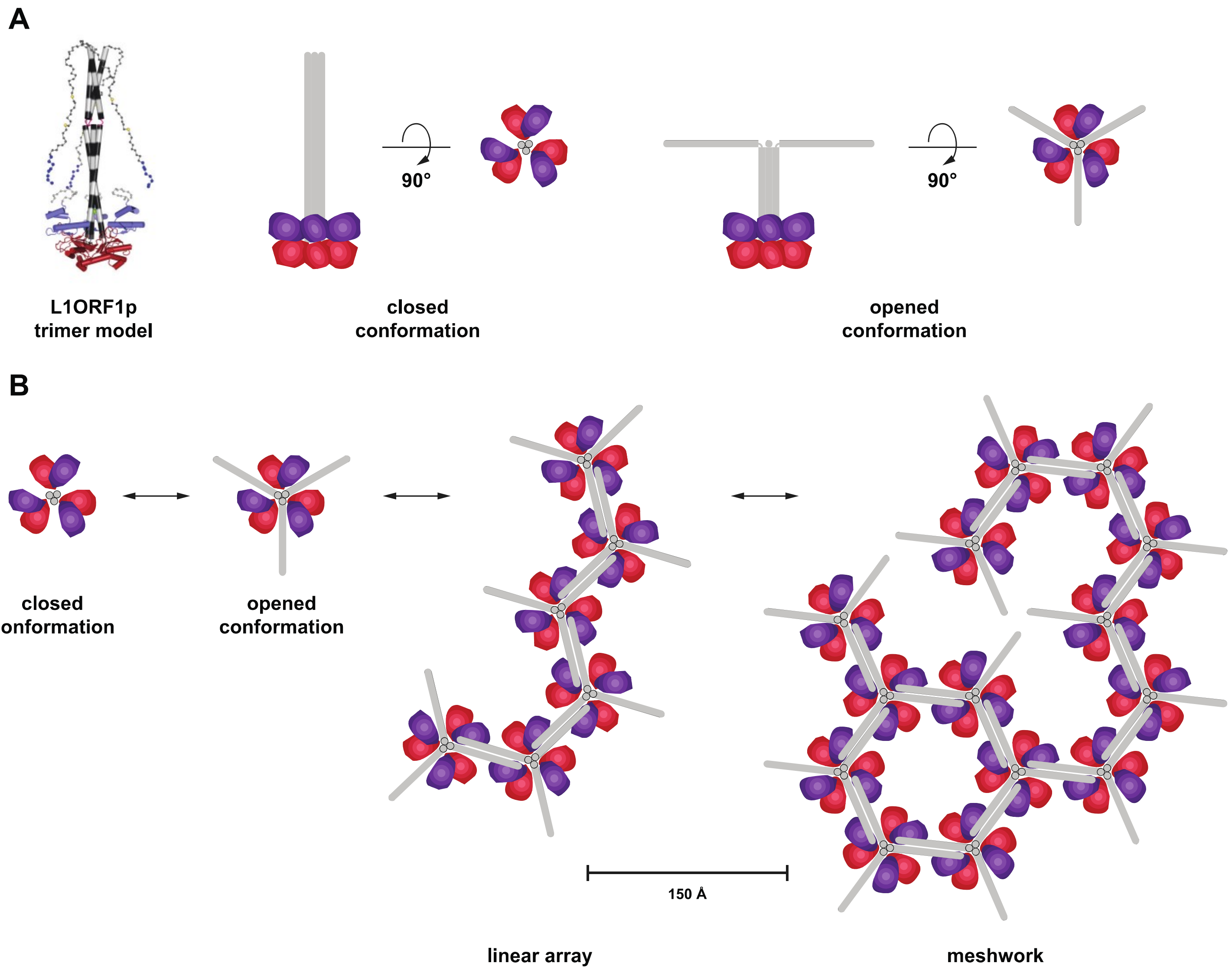

ORF1p
1
8
52
152
157
252
254
323
338
NTR
Coiled Coil
RRM
CTD
trialanine scanning library of L1 : 538 mutants
ORF1p :
1017 bp
338 residues
113 mutants
ORF2p :
3828 bp
1275 residues
425 mutants
ORF2p
1
239
347
380
480
498
773
1130
1148
1275
EN
Cry
Z
RT
Cys rich
D1
D2
Mutant 1
ORF1p residues 2-4
ATG GCT GCA GCT CAG AAC AGA AAA ACT GGA AAC TCT AAA ACG CAG AGC ...
M A A A Q N R K T G N S K T Q S ...
Mutant 2
ORF1p residues 5-7
ATG GGG AAA AAA GCT GCA GCT AAA ACT GGA AAC TCT AAA ACG CAG AGC ...
M G K K A A A K T G N S K T Q S ...
Mutant 538
ORF2p residues 1273-1275
... GAA CAA AAA ACC AAA CAC CGC ATA TTC TCA CTC ATA GCT GCA GCT TAG
... E Q K T K H R I F S L I A A A *

Figure 3.2 : The stammer in the coiled-coil of ORF1p introduces a flexibility that has inspired a theory for the formation of higher order structures by ORF1p homotrimers.
This figure is taken from Khazina and Weichenrieder, eLife (2018). (A) A simplified schematic is presented for. The ORF1p trimer : no NTR, the coiled coil In gray, the RRM in red and the CTD in blue. (B) Based on their structural and functional work, they propose that the flexibility of ORF1p homotrimers is essential and that the stammer allows for sampling of an "open" conformation that may induce formation of higher order "linear array" and "meshwork" structures.

See next page.


A
L1ORF1p
trimer model
90°
closed
conformation
90°
opened
conformation
B
closed
conformation
opened
conformation
150 Å
linear array
meshwork

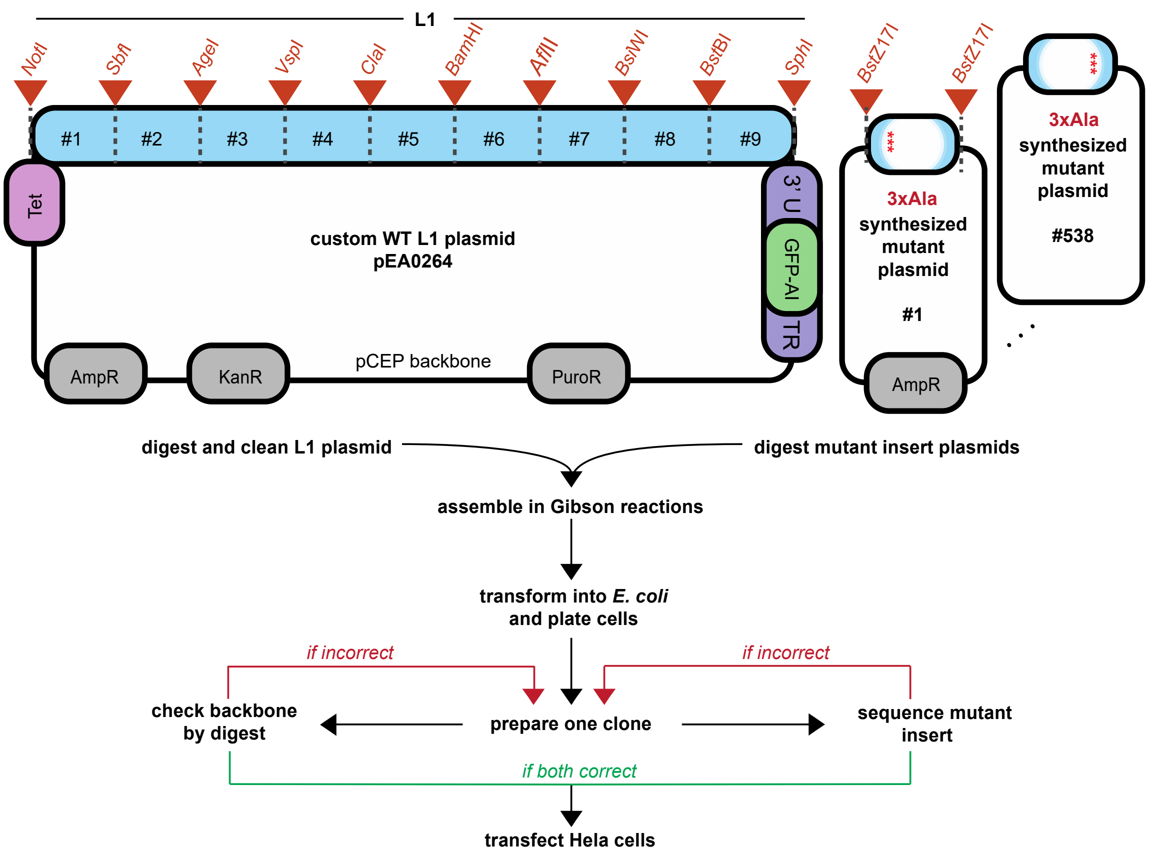


Figure 3.3 : The customized L1 construct and the build of the 538 trialanine variants.
In the upper left, the parental L1 plasmid, pEA0264 is diagrammed, featuring the engineered restriction sites. Orange triangles annotate the edges (designed unique restriction sites) of nine chunks. In the upper right, each of the 538 synthesized mutant plasmids were identical, excepting the 3xAla ~600bp fragment provided between the *BstZ17I* restriction sites. The pipeline for building the library is outlined below the plasmid schematics. An efficient two-piece Gibson assembly approach, followed by a two-part quality control procedure was used to build each mutant L1 construct in the library.

chunk #	range of mutated residues		total # residues	total # 3xAla mutants	unique flanking restriction sites	
	ORF1	ORF2				
1	2 - 175		174	58	*Not*I	*Sbf*I
2	176 - 328		153	51	*Sbf*I	*Age*I
3	329 - 338	2 - 144	153	52*	*Age*I	*Vsp*I
4		145 - 315	171	57	*Vsp*I	*Cla*I
5		316 - 504	189	63	*Cla*I	*Bam*HI
6		505 - 717	213	71	*Bam*HI	*Afl*II
7		718 - 903	186	62	*Afl*II	*Bsi*WI
8		904 - 1083	180	60	*Bsi*WI	*Bst*BI
9		1084 - 1275	192	64	*Bst*BI	*Sph*I

* The final variant for ORF1 mutates residue 338 (1xAla).
The first variant for ORF2 mutates residues 2-3 (2xAla).

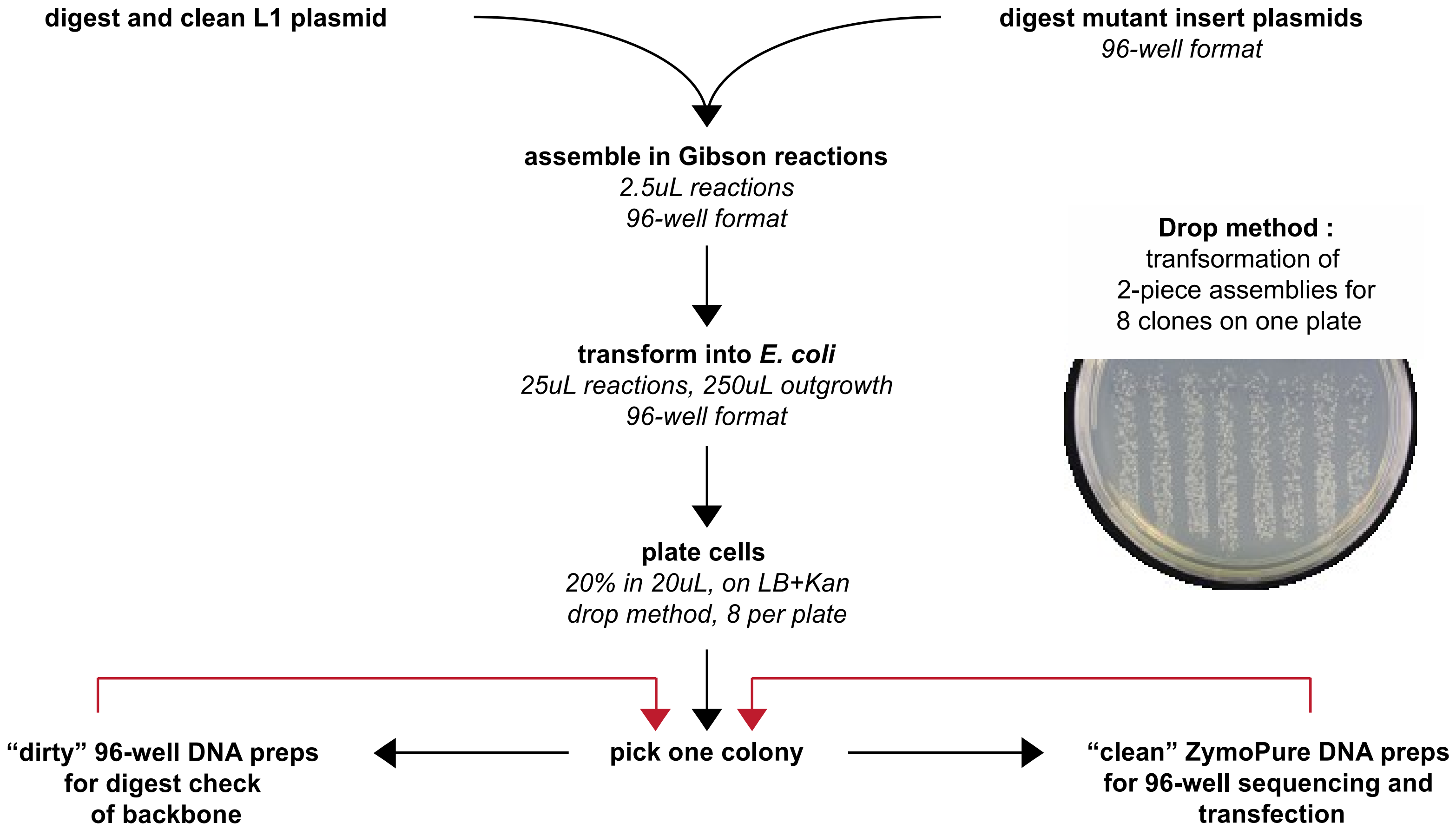
Table 3.1 : Organization of cloning trialanine variants into nine chunks
The L1 coding sequence was divided into nine chunks. Which mutants were contained in which chunk is shown in detail. This Table details the 338,886bp of DNA synthesis used to make the mutant fragments.

Plasmid ID (pEA####)	**Description**	**Notes**
0270 - 0806	synthesized DNA in plasmids	
0264	WT parental plasmid for trialanine library	(see Figure 3.3)
0807 - 1343	trialanine library, validated, in pEA0264 parental backbone	pEA919 has ORF1p residue 338 + ORF2p residue 2-3 mutated to alanine (it was not used in final presented data)
1348	adds same 9bp 3xala DNA sequence after the ORF2 stop codon in pEA0264	used as WT reference in sequencing experiments (retrotransposition frequency 120% that of pEA0264)
1361	mutates only residue 338 of ORF1p, in pEA0264 backbone	resolves mutating both proteins at once in pEA0919; it was used for final presented data
1362	mutates only residues 2-3 of ORF2p, in pEA0264 backbone	resolves mutating both proteins at once in pEA0919; it was used for final presented data
1440	pEA0264 without 4H1 anti-ORF1 ab antigen	control used in IPs of library (retrotransposition frequency 29% that of pEA0264)
1441	pEA0264 with H1 anti-ORF1 ab antigen mutated to alanines	control used in IPs of library (retrotransposition 76% frequency that of pEA0264)

Table 3.2 : Plasmids made in work with trialanine library.
This is an outline of the plasmids that were a part of the library build strategy. This includes the synthesized fragments that came as plasmids (from Gen9 or Qinglan) and the final library collection. See Boeke Bacterial Stock Collection for more details.

Figure 3.4 : Detailed high-throughput cloning procedure used to create library.
The detailed steps for building and validating the library in a high-throughput manner are depicted. The inset shows an agar plate after an overnight growth of eight transformations of putative trialanine clones after using the "drop method".

See next page.



digest and clean L1 plasmid

digest mutant insert plasmids
96-well format

assemble in Gibson reactions
2.5uL reactions
96-well format

transform into *E. coli*
25uL reactions, 250uL outgrowth
96-well format

plate cells
20% in 20uL, on LB+Kan
drop method, 8 per plate

pick one colony

“dirty” 96-well DNA preps
for digest check
of backbone

“clean” ZymoPure DNA preps
for 96-well sequencing and
transfection

Drop method :
tranfsormation of
2-piece assemblies for
8 clones on one plate

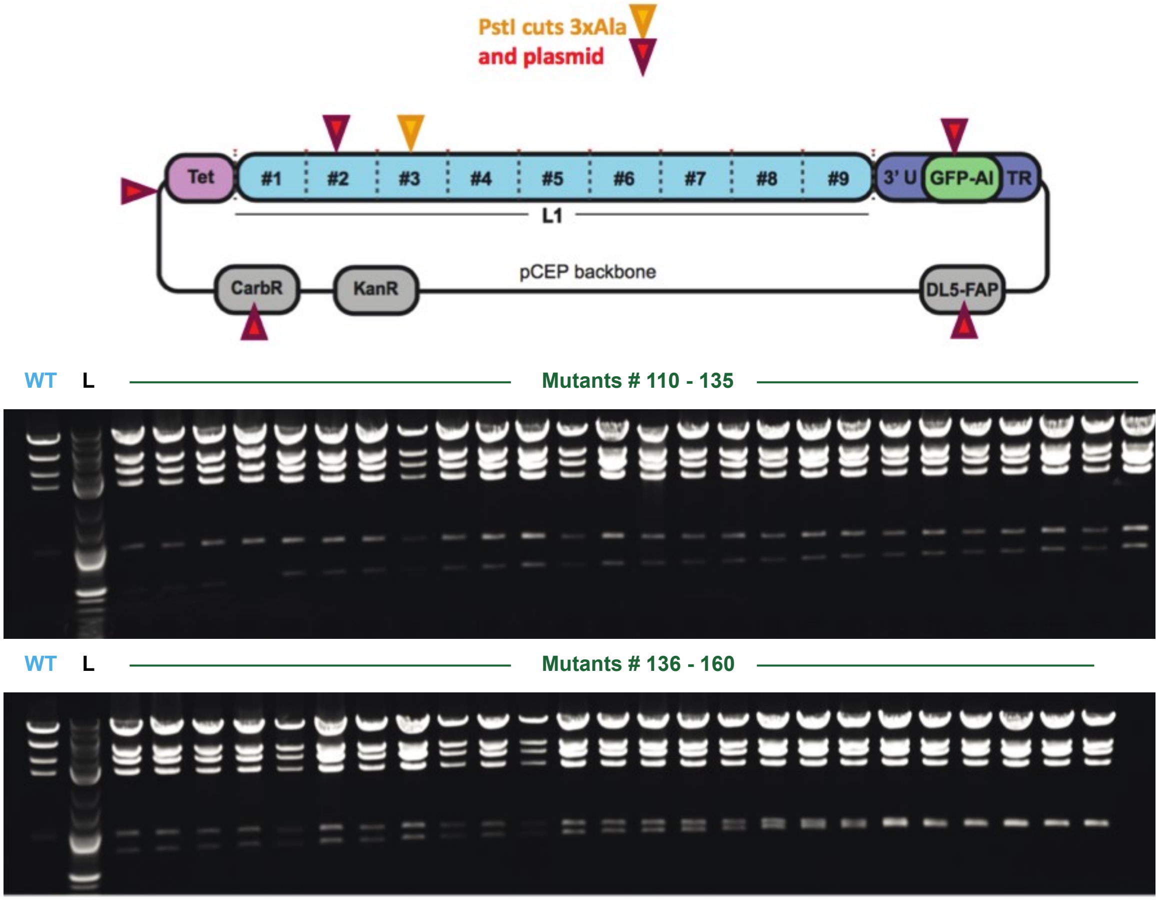


Figure 3.5 : Restriction digest used to validate final constructs in library. *Pst*I digestion was used to validate the final integrity of the backbone and the presence of the trialanine mutation for each clone. On top, a schematic of where *Pst*I cuts in the plasmid is shown. The trialanine DNA sequence was designed to contain a *Pst*I site, thus the predicted band sizes for each construct are unique. An agarose DNA gel shows diagnostic digests of each final correct clone in chunk 3, Mutants 110-160 ("WT" is the pEA0264 parent backbone and "L" is the 2-log ladder). Sanger sequencing (not shown) also confirmed the correct sequence for each mutant insert by spanning across the Gibson homology arm boundaries for each clone.

- ***Measurements of the retrotransposition frequency of each mutant***

We tested the ability of each variant to retrotranspose, since this is the most telling test for impaired function (values listed Table 3.3). Figure 3.6 shows the relative retrotransposition efficiency of each mutant and maps this value along the length of ORF1p and ORF2p, highlighting some key motifs and previously studied essential residues. See *Methods* and Figure 3.7 for details on the 96-well microscopy-based approach to these measurements and some controls done to prove the robustness and reproducibility of this technique in human cells.

About 50% of the trialanine mutants had a strong effect (defined as depleting retrotransposition activity to 25% or less that of WT), 34% had a mild effect (retrotransposition above 25% and below 80% that of WT), and only 16% retained wildtype activity (80%-125% of wild type). No mutants caused a significant increase in activity. Mutants containing ORF2p residues known from other studies to be critical for retrotransposition and thought to be catalytic (N13, E43, D145, D205, H230 and D702) all showed a strong effect with retrotransposition <20% of WT, providing a good calibration of the lowest activity category. By setting the threshold at 25% we allowed for some biological variation in any given mutant's retrotransposition level between experiments. A significant fraction of mutants (25% of ORF1p and 12% of ORF2p variants) had activity <=5% of WT, a much more stringent cutoff.

The proportions for the three categories of mutant impact described above (i.e., strong, mild and comparable to WT) were fairly consistent between ORF1p and ORF2p, with obvious clusters of strong effect in the more conserved domains of the proteins (Table 3.4). Mutation of well-conserved residues usually showed severely impaired retrotransposition. We also observed strong concordance with mutants affecting previously characterized residues. Overlaying the retrotransposition levels of the trialanine variants on the solved WT crystal structures gives a visual representation of each mutant's impact, for example the EN domain of ORF2p (Figure 3.8) The mapping of mutant phenotypes onto the full-length ORF1p structure will be presented visually in the next section.

Table 3.3 : Raw retrotransposition efficiency data for full library.
For ORF1p and ORF2p, this table shows the IDs and raw retrotransposition data for each mutant. The column *Trialanine construct ID* describes each mutant using the following format (with the information, separated by underscores) : the pEA construct number, the first residue mutated, the last residue mutated, and the three WT amino acids (single letter code) mutated to alanine. The second column shows the raw data that corresponds to Figure 2 : *retroT average*. The third and fourth columns show the *standard deviation* and *number of measurements*, respectively, for the measurements made for each mutant.

Table 3.3 spans the next five pages.

ORF1

Mutant ID	retroT average	standard deviation	number of measure-ments	Mutant ID	retroT average	standard deviation	number of measure-ments
pEA807_2_4_GKK	4.8	3.8	4	pEA863_170_172_NGT	74.5	11.8	4
pEA808_5_7_QNR	76.2	3.4	4	pEA864_173_175_KLE	20.3	2.8	2
pEA809_8_10_KTG	111.0	7.1	4	pEA865_176_178_NTL	4.0	2.6	4
pEA810_11_13_NSK	95.5	27.3	2	pEA866_179_181_QDI	1.6	1.9	4
pEA811_14_16_TQS	85.9	14.0	4	pEA867_182_184_IQE	6.6	9.8	4
pEA812_17_19_ASP	82.0	29.4	4	pEA868_185_187_NFP	5.8	9.1	4
pEA813_20_22_PPK	83.2	16.2	4	pEA869_188_190_NLA	51.7	12.3	2
pEA814_23_25_ERS	70.2	17.8	4	pEA870_191_193_RQA	101.2	22.8	4
pEA815_26_28_SSP	73.1	13.3	4	pEA871_194_196_NVQ	92.7	9.9	4
pEA816_29_31_ATE	100.4	24.6	4	pEA872_197_199_IQE	30.0	5.4	2
pEA817_32_34_QSW	81.4	17.3	4	pEA873_200_202_IQR	2.5	3.3	4
pEA818_35_37_MEN	93.1	14.3	6	pEA874_203_205_TPQ	5.6	1.8	4
pEA819_38_40_DFD	95.9	14.3	6	pEA875_206_208_RYS	6.9	6.0	4
pEA820_41_43_ELR	97.3	13.3	6	pEA876_209_211_SRR	9.2	4.3	4
pEA821_44_46_EEG	58.5	18.6	4	pEA877_212_214_ATP	10.9	4.9	4
pEA822_47_49_FRR	76.1	4.5	2	pEA878_215_217_RHI	2.7	3.5	4
pEA823_50_52_SNY	53.8	19.4	2	pEA879_218_220_IVR	5.3	1.2	2
pEA824_53_55_SEL	9.0	10.3	4	pEA880_221_223_FTK	11.5	2.8	2
pEA825_56_58_RED	72.5	8.0	2	pEA881_224_226_VEM	53.4	17.6	4
pEA826_59_61_IQT	10.2	4.9	4	pEA882_227_229_KEK	49.5	14.6	2
pEA827_62_64_KGK	78.0	9.3	2	pEA883_230_232_MLR	4.3	6.2	4
pEA828_65_67_EVE	8.4	4.5	4	pEA884_233_235_AAR	36.6	13.4	4
pEA829_68_70_NFE	12.9	10.9	4	pEA885_236_238_EKG	76.8	8.4	4
pEA830_71_73_KNL	25.1	9.9	4	pEA886_239_241_RVT	4.1	2.8	4
pEA831_74_76_EEC	68.9	11.3	4	pEA887_242_244_LKG	4.2	4.4	4
pEA832_77_79_ITR	8.8	4.7	2	pEA888_245_247_KPI	3.9	3.5	4
pEA833_80_82_ITN	42.2	13.1	3	pEA889_248_250_RLT	5.2	8.0	4
pEA834_83_85_TEK	8.3	7.2	3	pEA890_251_253_ADL	3.4	3.4	4
pEA835_86_88_CLK	9.8	6.0	4	pEA891_254_256_SAE	68.8	11.0	4
pEA836_89_91_ELM	4.2	1.3	4	pEA892_257_259_TLQ	5.6	3.1	4
pEA837_92_94_ELK	0.7	1.0	2	pEA893_260_262_ARR	3.2	3.0	4
pEA838_95_97_TKA	103.1	10.5	4	pEA894_263_265_EWG	4.0	1.6	4
pEA839_98_100_REL	5.8	3.0	4	pEA895_266_268_PIF	2.5	2.5	4
pEA840_101_103_REE	33.0	7.7	4	pEA896_269_271_NIL	2.5	1.9	4
pEA841_104_106_CRS	7.7	0.5	4	pEA897_272_274_KEK	57.6	11.2	4
pEA842_107_109_LRS	9.4	4.6	4	pEA898_275_277_NFQ	13.2	1.3	4
pEA843_110_112_RCD	10.1	5.1	4	pEA899_278_280_PRI	3.4	3.5	4
pEA844_113_115_QLE	2.9	2.1	4	pEA900_281_283_SYP	5.9	3.9	4
pEA845_116_118_ERV	13.1	11.2	4	pEA901_284_286_AKL	2.8	2.6	4
pEA846_119_121_SAM	36.4	2.8	4	pEA902_287_289_SFI	4.0	5.3	4
pEA847_122_124_EDE	4.0	1.8	4	pEA903_290_292_SEG	3.0	2.8	4
pEA848_125_127_MNE	60.0	3.1	4	pEA904_293_295_EIK	63.0	20.9	4
pEA849_128_130_MKR	34.8	6.4	4	pEA905_296_298_YFI	2.0	1.7	4
pEA850_131_133_EGK	91.4	16.4	4	pEA906_299_301_DKQ	1.9	1.9	4
pEA851_134_136_FRE	86.5	10.5	4	pEA907_302_304_MLR	24.5	8.4	4
pEA852_137_139_KRI	50.8	5.2	4	pEA908_305_307_DFV	4.7	1.1	2
pEA853_140_142_KRN	35.5	7.8	4	pEA909_308_310_TTR	60.6	13.2	4
pEA854_143_145_EQS	63.2	11.4	4	pEA910_311_313_PAL	2.4	2.8	4
pEA855_146_148_LQE	13.9	5.4	3	pEA911_314_316_KEL	79.3	17.4	2
pEA856_149_151_IWD	5.0	6.2	4	pEA912_317_319_LKE	13.4	3.1	4
pEA857_152_154_YVK	3.9	2.2	4	pEA913_320_322_ALN	33.5	15.0	4
pEA858_155_157_RPN	7.1	5.9	4	pEA914_323_325_MER	37.9	9.6	4
pEA859_158_160_LRL	5.9	4.9	4	pEA915_326_328_NNR	76.5	22.1	4
pEA860_161_163_IGV	11.0	9.6	3	pEA916_329_331_YQP	107.3	14.5	6
pEA861_164_166_PES	11.4	1.3	2	pEA917_332_334_LQN	84.6	20.1	4
pEA862_167_169_DVE	45.8	0.3	2	pEA918_335_337_HAK	85.1	17.0	4
				pEA1361_338_M	77.4	12.2	4

ORF2

Mutant ID	retroT average	standard deviation	number of measure-ments	Mutant ID	retroT average	standard deviation	number of measure-ments
pEA1362_2_3_TG	24.8	1.5	4	pEA981_187_189_KST	87.9	11.6	4
pEA920_4_6_STS	72.0	5.5	2	pEA982_190_192_EYT	3.9	2.8	4
pEA921_7_9_HIT	17.6	20.4	4	pEA983_193_195_FFS	2.4	2.2	4
pEA922_10_12_ILT	4.3	5.0	4	pEA984_196_198_APH	10.3	5.2	4
pEA923_13_15_LNI	6.3	4.3	4	pEA985_199_201_HTY	6.0	1.1	2
pEA924_16_18_NGL	4.8	3.0	4	pEA986_202_204_SKI	7.5	5.1	4
pEA925_19_21_NSA	8.6	3.4	4	pEA987_205_207_DHI	1.9	1.8	4
pEA926_22_24_IKR	10.7	12.9	4	pEA988_208_210_VGS	6.6	5.1	4
pEA927_25_27_HRL	36.5	8.7	3	pEA989_211_213_KAL	6.0	3.4	4
pEA928_28_30_ASW	18.3	2.6	4	pEA990_214_216_LSK	61.5	23.9	4
pEA929_31_33_IKS	10.0	11.6	4	pEA991_217_219_CKR	91.6	29.0	4
pEA930_34_36_QDP	42.7	6.0	2	pEA992_220_222_TEI	45.9	8.8	4
pEA931_37_39_SVC	0.0	0.0	2	pEA993_223_225_ITN	33.4	8.6	4
pEA932_40_42_CIQ	2.3	2.5	4	pEA994_226_228_YLS	16.8	5.1	2
pEA933_43_45_ETH	1.7	2.1	4	pEA995_229_231_DHS	2.9	3.0	4
pEA934_46_48_LTC	5.8	3.9	4	pEA996_232_234_AIK	13.7	7.7	4
pEA935_49_51_RDT	81.1	3.7	4	pEA997_235_237_LEL	7.3	1.8	2
pEA936_52_54_HRL	4.2	5.0	4	pEA998_238_240_RIK	91.3	28.2	4
pEA937_55_57_KIK	15.0	7.0	4	pEA999_241_243_NLT	103.9	28.5	6
pEA938_58_60_GWR	3.9	0.8	2	pEA1000_244_246_QSR	60.4	13.1	2
pEA939_61_63_KIY	0.0	0.0	2	pEA1001_247_249_STT	87.9	26.3	4
pEA940_64_66_QAN	83.8	20.3	4	pEA1002_250_252_WKL	4.5	4.8	4
pEA941_67_69_GKQ	52.0	11.5	4	pEA1003_253_255_NNL	41.7	9.8	4
pEA942_70_72_KKA	7.2	4.6	4	pEA1004_256_258_LLN	9.6	2.0	4
pEA943_73_75_GVA	3.5	3.5	4	pEA1005_259_261_DYW	29.4	15.9	4
pEA944_76_78_ILV	1.6	1.8	4	pEA1006_262_264_VHN	106.1	12.5	2
pEA945_79_81_SDK	2.2	1.3	4	pEA1007_265_267_EMK	21.5	7.7	4
pEA946_82_84_TDF	6.3	0.1	2	pEA1008_268_270_AEI	13.0	9.9	3
pEA947_85_87_KPT	17.8	20.7	4	pEA1009_271_273_KMF	7.3	6.3	4
pEA948_88_90_KIK	10.0	2.3	4	pEA1010_274_276_FET	16.2	2.7	2
pEA949_91_93_RDK	10.0	7.6	2	pEA1011_277_279_NEN	12.2	6.9	4
pEA950_94_96_EGH	3.1	2.0	4	pEA1012_280_282_KDT	27.6	6.4	4
pEA951_97_99_YIM	2.3	2.5	4	pEA1013_283_285_TYQ	18.6	11.7	4
pEA952_100_102_VKG	3.4	1.8	4	pEA1014_286_288_NLW	2.6	2.7	4
pEA953_103_105_SIQ	27.8	8.3	4	pEA1015_289_291_DAF	3.2	2.2	4
pEA954_106_108_QEE	32.9	0.6	2	pEA1016_292_294_KAV	2.5	2.8	4
pEA955_109_111_LTI	5.3	4.6	3	pEA1017_295_297_CRG	4.8	4.2	4
pEA956_112_114_LNI	2.5	2.9	4	pEA1018_298_300_KFI	3.0	3.2	4
pEA957_115_117_YAP	2.2	1.5	4	pEA1019_301_303_ALN	103.0	25.1	4
pEA958_118_120_NTG	24.1	7.9	4	pEA1020_304_306_AYK	73.5	18.6	4
pEA959_121_123_APR	65.5	4.5	2	pEA1021_307_309_RKQ	62.5	4.2	2
pEA960_124_126_FIK	4.8	3.6	4	pEA1022_310_312_ERS	64.3	16.7	4
pEA961_127_129_QVL	2.1	2.1	4	pEA1023_313_315_KID	62.2	20.1	4
pEA962_130_132_SDL	39.8	3.6	2	pEA1024_316_318_TLT	59.4	5.4	4
pEA963_133_135_QRD	23.4	27.0	4	pEA1025_319_321_SQL	72.9	13.9	3
pEA964_136_138_LDS	7.2	0.1	2	pEA1026_322_324_KEL	68.7	14.0	4
pEA965_139_141_HTL	2.7	3.1	4	pEA1027_325_327_EKQ	86.9	7.9	4
pEA966_142_144_IMG	2.4	2.4	4	pEA1028_328_330_EQT	76.7	18.9	4
pEA967_145_147_DFN	4.0	4.6	4	pEA1029_331_333_HSK	82.8	20.4	4
pEA968_148_150_TPL	6.2	3.5	4	pEA1030_334_336_ASR	93.7	23.7	4
pEA969_151_153_STL	63.8	21.4	4	pEA1031_337_339_RQE	19.6	9.4	4
pEA970_154_156_DRS	2.5	2.9	4	pEA1032_340_342_ITK	31.1	3.5	4
pEA971_157_159_TRQ	51.5	15.9	4	pEA1033_343_345_IRA	29.6	8.2	4
pEA972_160_162_KVN	48.1	16.4	4	pEA1034_346_348_ELK	14.2	9.9	4
pEA973_163_165_KDT	96.8	11.9	4	pEA1035_349_351_EIE	59.0	8.5	2
pEA974_166_168_QEL	19.1	3.1	4	pEA1036_352_354_TQK	53.5	8.1	4
pEA975_169_171_NSA	79.4	21.8	4	pEA1037_355_357_TLQ	30.7	2.8	4
pEA976_172_174_LHQ	25.0	14.2	4	pEA1038_358_360_KIN	56.8	31.2	4
pEA977_175_177_ADL	8.4	0.7	2	pEA1039_361_363_ESR	31.5	15.2	3
pEA978_178_180_IDI	2.8	3.2	4	pEA1040_364_366_SWF	11.7	1.9	3
pEA979_181_183_YRT	2.9	2.1	4	pEA1041_367_369_FER	23.6	1.3	2
pEA980_184_186_LHP	4.9	0.7	2	pEA1042_370_372_INK	17.3	7.1	4

ORF2

Mutant ID	retroT average	standard deviation	number of measure-ments
pEA1043_373_375_IDR	61.2	11.0	4
pEA1044_376_378_PLA	41.7	13.5	4
pEA1045_379_381_RLI	65.1	8.6	4
pEA1046_382_384_KKK	56.9	0.4	2
pEA1047_385_387_REK	58.6	1.3	2
pEA1048_388_390_NQI	59.8	20.4	4
pEA1049_391_393_DTI	60.3	12.5	6
pEA1050_394_396_KND	76.5	33.4	4
pEA1051_397_399_KGD	55.1	24.2	4
pEA1052_400_402_ITT	22.5	7.0	4
pEA1053_403_405_DPT	45.1	7.8	4
pEA1054_406_408_EIQ	14.4	5.0	4
pEA1055_409_411_TTI	27.7	1.0	2
pEA1056_412_414_REY	27.9	3.8	3
pEA1057_415_417_YKH	33.2	6.6	4
pEA1058_418_420_LYA	30.7	6.0	4
pEA1059_421_423_NKL	65.6	10.7	4
pEA1060_424_426_ENL	49.6	7.7	4
pEA1061_427_429_EEM	38.0	9.4	4
pEA1062_430_432_DTF	20.1	20.2	4
pEA1063_433_435_LDT	5.5	4.4	4
pEA1064_436_438_YTL	18.5	14.5	4
pEA1065_439_441_PRL	22.6	34.5	4
pEA1066_442_444_NQE	73.3	11.8	4
pEA1067_445_447_EVE	42.9	16.2	4
pEA1068_448_450_SLN	30.3	12.9	3
pEA1069_451_453_RPI	30.8	33.2	4
pEA1070_454_456_TGS	87.5	18.3	4
pEA1071_457_459_EIV	3.4	2.5	4
pEA1072_460_462_AII	16.3	4.1	3
pEA1073_463_465_NSL	12.6	12.9	3
pEA1074_466_468_PTK	40.0	4.9	2
pEA1075_469_471_KSP	61.9	13.8	2
pEA1076_472_474_GPD	3.5	2.2	4
pEA1077_475_477_GFT	7.4	6.9	4
pEA1078_478_480_AEF	18.5	10.6	4
pEA1079_481_483_YQR	14.4	13.8	4
pEA1080_484_486_YKE	8.0	2.4	2
pEA1081_487_489_ELV	9.7	6.4	4
pEA1082_490_492_PFL	8.1	7.5	4
pEA1083_493_495_LKL	6.2	2.0	4
pEA1084_496_498_FQS	28.1	8.4	4
pEA1085_499_501_IEK	79.3	27.2	4
pEA1086_502_504_EGI	38.1	9.5	4
pEA1087_505_507_LPN	8.2	1.6	4
pEA1088_508_510_SFY	8.9	2.4	4
pEA1089_511_513_EAS	41.1	6.5	4
pEA1090_514_516_IIL	5.9	2.3	4
pEA1091_517_519_IPK	8.2	0.9	3
pEA1092_520_522_PGR	9.9	6.1	4
pEA1093_523_525_DTT	80.1	24.7	4
pEA1094_526_528_KKE	82.6	6.7	4
pEA1095_529_531_NFR	5.6	0.7	4
pEA1096_532_534_PIS	5.7	2.6	4
pEA1097_535_537_LMN	9.7	5.4	2
pEA1098_538_540_IDA	8.0	6.7	4
pEA1099_541_543_KIL	4.5	1.5	4
pEA1100_544_546_NKI	58.1	7.5	4
pEA1101_547_549_LAN	53.5	3.8	4
pEA1102_550_552_RIQ	8.1	2.3	4
pEA1103_553_555_QHI	88.6	17.3	4
pEA1104_556_558_KKL	43.4	4.8	4
pEA1105_559_561_IHH	16.3	11.5	4
pEA1106_562_564_DQV	6.2	3.5	4
pEA1107_565_567_GFI	20.4	7.8	4
pEA1108_568_570_PGM	53.3	9.2	4
pEA1109_571_573_QGW	15.8	3.1	4
pEA1110_574_576_FNI	36.0	12.5	4
pEA1111_577_579_RKS	4.7	1.1	4
pEA1112_580_582_INV	30.0	8.2	4
pEA1113_583_585_IQH	61.4	4.7	4
pEA1114_586_588_INR	57.4	17.8	4
pEA1115_589_591_AKD	78.6	11.3	4
pEA1116_592_594_KNH	53.5	8.9	4
pEA1117_595_597_MII	6.3	2.0	4
pEA1118_598_600_SID	5.3	1.3	4
pEA1119_601_603_AEK	62.1	9.5	4
pEA1120_604_606_AFD	3.8	1.5	4
pEA1121_607_609_KIQ	38.7	6.4	4
pEA1122_610_612_QPF	5.1	0.6	2
pEA1123_613_615_MLK	6.4	1.9	4
pEA1124_616_618_TLN	35.5	3.2	4
pEA1125_619_621_KLG	13.0	2.1	3
pEA1126_622_624_IDG	7.9	4.3	4
pEA1127_625_627_TYF	18.8	7.9	4
pEA1128_628_630_KII	4.3	1.4	4
pEA1129_631_633_RAI	25.4	6.2	4
pEA1130_634_636_YDK	13.6	1.0	4
pEA1131_637_639_PTA	62.5	12.6	4
pEA1132_640_642_NII	4.1	0.8	4
pEA1133_643_645_LNG	13.0	2.3	4
pEA1134_646_648_QKL	44.5	4.6	4
pEA1135_649_651_EAF	6.2	3.5	4
pEA1136_652_654_PLK	35.3	5.1	4
pEA1137_655_657_TGT	10.7	1.1	4
pEA1138_658_660_RQG	8.7	5.6	4
pEA1139_661_663_CPL	5.6	1.5	4
pEA1140_664_666_SPL	4.8	1.3	4
pEA1141_667_669_LFN	11.6	3.0	4
pEA1142_670_672_IVL	13.9	5.4	4
pEA1143_673_675_EVL	4.9	1.2	4
pEA1144_676_678_ARA	123.2	15.7	4
pEA1145_679_681_IRQ	5.9	1.0	4
pEA1146_682_684_EKE	29.2	5.8	4
pEA1147_685_687_IKG	4.4	1.4	4
pEA1148_688_690_IQL	9.1	1.9	4
pEA1149_691_693_GKE	51.7	12.3	4
pEA1150_694_696_EVK	3.4	1.1	4
pEA1151_697_699_LSL	36.9	4.3	4
pEA1152_700_702_FAD	9.4	6.3	4
pEA1153_703_705_DMI	4.0	1.3	4
pEA1154_706_708_VYL	7.3	3.0	3
pEA1155_709_711_ENP	15.8	7.2	3
pEA1156_712_714_IVS	34.3	4.1	4
pEA1157_715_717_AQN	88.0	12.9	3
pEA1158_718_720_LLK	7.3	4.0	4

ORF2

Mutant ID	retroT average	standard deviation	number of measure-ments	Mutant ID	retroT average	standard deviation	number of measure-ments
pEA1159_721_723_LIS	6.0	3.1	4	pEA1221_907_909_IMP	112.1	4.0	4
pEA1160_724_726_NFS	5.4	1.1	4	pEA1222_910_912_HIY	88.5	4.0	4
pEA1161_727_729_KVS	20.1	7.2	4	pEA1223_913_915_NYL	105.1	4.0	4
pEA1162_730_732_GYK	5.6	2.6	4	pEA1224_916_918_IFD	7.3	4.0	4
pEA1163_733_735_INV	11.4	3.2	4	pEA1225_919_921_KPE	95.3	4.0	4
pEA1164_736_738_QKS	5.3	1.9	4	pEA1226_922_924_KNK	79.0	4.0	4
pEA1165_739_741_QAF	8.5	2.9	4	pEA1227_925_927_QWG	112.0	4.0	4
pEA1166_742_744_LYT	6.4	2.2	4	pEA1228_928_930_KDS	73.2	4.0	4
pEA1167_745_747_NNR	59.7	10.0	4	pEA1229_931_933_LFN	48.9	4.0	4
pEA1168_748_750_QTE	92.1	7.4	4	pEA1230_934_936_KWC	75.3	4.0	4
pEA1169_751_753_SQI	119.7	4.0	4	pEA1231_937_939_WEN	64.5	4.0	4
pEA1170_754_756_MGE	115.8	11.8	6	pEA1232_940_942_WLA	18.7	4.0	4
pEA1171_757_759_LPF	22.2	0.9	4	pEA1233_943_945_ICR	96.0	4.0	4
pEA1172_760_762_TIA	92.5	27.1	4	pEA1234_946_948_KLK	97.5	4.0	4
pEA1173_763_765_SKR	66.0	14.7	4	pEA1235_949_951_LDP	70.9	4.0	4
pEA1174_766_768_IKY	10.2	3.2	4	pEA1236_952_954_FLT	16.2	4.0	4
pEA1175_769_771_LGI	4.3	2.4	4	pEA1237_955_957_PYT	3.8	4.0	4
pEA1176_772_774_QLT	20.0	3.0	4	pEA1238_958_960_KIN	7.9	4.0	4
pEA1177_775_777_RDV	78.3	17.3	4	pEA1239_961_963_SRW	9.7	4.0	4
pEA1178_778_780_KDL	44.3	11.3	4	pEA1240_964_966_IKD	42.0	4.0	4
pEA1179_781_783_FKE	82.7	8.9	4	pEA1241_967_969_LNV	7.6	4.0	4
pEA1180_784_786_NYK	14.2	2.7	4	pEA1242_970_972_KPK	64.2	4.0	4
pEA1181_787_789_PLL	42.3	2.2	4	pEA1243_973_975_TIK	13.8	2.0	2
pEA1182_790_792_KEI	53.8	5.4	4	pEA1244_976_978_TLE	27.6	4.0	4
pEA1183_793_795_KEE	75.4	12.4	4	pEA1245_979_981_ENL	63.0	4.0	4
pEA1184_796_798_TNK	104.1	15.3	4	pEA1246_982_984_GIT	25.0	3.0	3
pEA1185_799_801_WKN	62.1	6.6	4	pEA1247_985_987_IQD	50.7	4.0	4
pEA1186_802_804_IPC	79.9	9.3	4	pEA1248_988_990_IGV	10.2	4.0	4
pEA1187_805_807_SWV	24.5	9.2	4	pEA1249_991_993_GKD	97.4	4.0	4
pEA1188_808_810_GRI	34.3	4.9	4	pEA1250_994_996_FMS	8.0	4.0	4
pEA1189_811_813_NIV	41.4	7.7	4	pEA1251_997_999_KTP	93.4	4.0	4
pEA1190_814_816_KMA	35.3	15.8	4	pEA1252_1000_1002_KAM	103.2	4.0	4
pEA1191_817_819_ILP	10.7	3.3	4	pEA1253_1003_1005_ATK	108.8	4.0	4
pEA1192_820_822_KVI	19.4	3.8	4	pEA1254_1006_1008_DKI	77.3	4.0	4
pEA1193_823_825_YRF	100.1	26.0	4	pEA1255_1009_1011_DKW	11.0	4.0	4
pEA1194_826_828_NAI	55.8	3.9	4	pEA1256_1012_1014_DLI	10.6	4.0	4
pEA1195_829_831_PIK	5.2	0.4	4	pEA1257_1015_1017_KLK	10.7	4.0	4
pEA1196_832_834_LPM	66.1	21.2	4	pEA1258_1018_1020_SFC	6.4	4.0	4
pEA1197_835_837_TFF	34.8	17.9	4	pEA1259_1021_1023_TAK	92.8	4.0	4
pEA1198_838_840_TEL	67.4	13.5	4	pEA1260_1024_1026_ETT	91.0	4.0	4
pEA1199_841_843_EKT	82.5	6.4	4	pEA1261_1027_1029_IRV	109.8	4.0	4
pEA1200_844_846_TLK	97.6	18.3	2	pEA1262_1030_1032_NRQ	71.9	3.0	3
pEA1201_847_849_FIW	8.2	2.4	4	pEA1263_1033_1035_PTT	125.7	4.0	4
pEA1202_850_852_NQK	83.3	14.9	4	pEA1264_1036_1038_WEK	28.5	4.0	4
pEA1203_853_855_RAR	18.9	4.0	4	pEA1265_1039_1041_IFA	7.8	4.0	4
pEA1204_856_858_IAK	52.0	4.0	4	pEA1266_1042_1044_TYS	106.5	4.0	4
pEA1205_859_861_SIL	56.0	4.0	4	pEA1267_1045_1047_SDK	40.6	4.0	4
pEA1206_862_864_SQK	72.9	4.0	4	pEA1268_1048_1050_GLI	14.6	4.0	4
pEA1207_865_867_NKA	122.1	4.0	4	pEA1269_1051_1053_SRI	55.3	2.0	2
pEA1208_868_870_GGI	60.4	4.0	4	pEA1270_1054_1056_YNE	66.0	4.0	4
pEA1209_871_873_TLP	69.2	4.0	4	pEA1271_1057_1059_LKQ	82.3	4.0	4
pEA1210_874_876_DFK	6.7	4.0	4	pEA1272_1060_1062_IYK	66.5	2.0	2
pEA1211_877_879_LYY	5.6	4.0	4	pEA1273_1063_1065_KKT	107.5	4.0	4
pEA1212_880_882_KAT	59.9	4.0	4	pEA1274_1066_1068_NNP	56.6	4.0	4
pEA1213_883_885_VTK	62.3	4.0	4	pEA1275_1069_1071_IKK	42.6	2.0	2
pEA1214_886_888_TAW	36.4	4.0	4	pEA1276_1072_1074_WAK	17.6	4.0	4
pEA1215_889_891_YWY	3.9	4.0	4	pEA1277_1075_1077_DMN	5.4	4.0	4
pEA1216_892_894_QNR	37.8	4.0	4	pEA1278_1078_1080_RHF	14.2	4.0	4
pEA1217_895_897_DID	54.9	4.0	4	pEA1279_1081_1083_SKE	93.5	4.0	4
pEA1218_898_900_QWN	59.9	4.0	4	pEA1280_1084_1086_DIY	65.1	4.0	4
pEA1219_901_903_RTE	68.9	4.0	4	pEA1281_1087_1089_AAK	64.0	4.0	4
pEA1220_904_906_PSE	119.5	4.0	4	pEA1282_1090_1092_KHM	14.9	4.0	4

ORF2

Mutant ID	retroT average	standard deviation	number of measure-ments
pEA1283_1093_1095_KKC	65.4	4.0	4
pEA1284_1096_1098_SSS	43.6	4.0	4
pEA1285_1099_1101_LAI	50.2	4.0	4
pEA1286_1102_1104_REM	24.2	4.0	4
pEA1287_1105_1107_QIK	27.9	4.0	4
pEA1288_1108_1110_TTM	43.7	4.0	4
pEA1289_1111_1113_RYH	5.3	2.0	2
pEA1290_1114_1116_LTP	21.1	2.0	2
pEA1291_1117_1119_VRM	70.3	4.0	4
pEA1292_1120_1122_AII	116.5	6.0	6
pEA1293_1123_1125_KKS	92.7	4.0	4
pEA1294_1126_1128_GNN	97.2	4.0	4
pEA1295_1129_1131_RCW	7.9	4.0	4
pEA1296_1132_1134_RGC	6.3	4.0	4
pEA1297_1135_1137_GEI	126.9	4.0	4
pEA1298_1138_1140_GTL	82.0	4.0	4
pEA1299_1141_1143_LHC	5.2	4.0	4
pEA1300_1144_1146_WWD	18.7	4.0	4
pEA1301_1147_1149_CKL	11.3	4.0	4
pEA1302_1150_1152_VQP	46.3	4.0	4
pEA1303_1153_1155_LWK	10.9	4.0	4
pEA1304_1156_1158_SVW	23.6	4.0	4
pEA1305_1159_1161_RFL	87.2	4.0	4
pEA1306_1162_1164_RDL	100.4	4.0	4
pEA1307_1165_1167_ELE	79.5	4.0	4
pEA1308_1168_1170_IPF	49.7	4.0	4
pEA1309_1171_1173_DPA	69.7	4.0	4
pEA1310_1174_1176_IPL	21.3	4.0	4
pEA1311_1177_1179_LGI	4.5	4.0	4
pEA1312_1180_1182_YPN	86.7	4.0	4
pEA1313_1183_1185_EYK	109.3	4.0	4
pEA1314_1186_1188_SCC	97.2	5.0	5
pEA1315_1189_1191_YKD	94.0	4.0	4
pEA1316_1192_1194_TCT	63.9	4.0	4
pEA1317_1195_1197_RMF	19.2	4.0	4
pEA1318_1198_1200_IAA	102.4	6.0	6
pEA1319_1201_1203_LFT	82.2	4.0	4
pEA1320_1204_1206_IAK	68.5	2.0	2
pEA1321_1207_1209_TWN	13.1	2.0	2
pEA1322_1210_1212_QPK	85.6	2.0	2
pEA1323_1213_1215_CPT	63.0	2.0	2
pEA1324_1216_1218_MID	91.0	4.0	4
pEA1325_1219_1221_WIK	24.9	4.0	4
pEA1326_1222_1224_KMW	7.8	4.0	4
pEA1327_1225_1227_HIY	24.1	4.0	4
pEA1328_1228_1230_TME	4.6	4.0	4
pEA1329_1231_1233_YYA	23.3	2.0	2
pEA1330_1234_1236_AIK	50.8	2.0	2
pEA1331_1237_1239_NDE	61.1	3.0	3
pEA1332_1240_1242_FIS	59.6	4.0	4
pEA1333_1243_1245_FVG	28.4	3.0	3
pEA1334_1246_1248_TWM	16.3	4.0	4
pEA1335_1249_1251_KLE	20.6	2.0	2
pEA1336_1252_1254_TII	8.8	4.0	4
pEA1337_1255_1257_LSK	56.4	4.0	4
pEA1338_1258_1260_LSQ	82.5	3.0	3
pEA1339_1261_1263_EQK	104.1	6.0	6
pEA1340_1264_1266_TKH	101.8	8.0	8
pEA1341_1267_1269_RIF	48.8	4.0	4
pEA1342_1270_1272_SLI	80.8	4.0	4
pEA1343_1273_1275_GGN	58.4	8.0	8

Figure 3.6 : The retrotransposition efficiency of each trialanine variant. Along the top are the schematics of ORF1p and ORF2p, highlighting domain boundaries as well as well characterized motifs and essential residues. The residue position is indicated along the x-axis and the y-axis denotes the percentage of WT activity each variant had. Each trialanine variant's average retrotransposition efficiency is graphed. Each mutant's retrotransposition has been normalized to WT measurements made in the same experiment on the same plate. WT retrotransposes at 100%, as the gray bar shows. Statistically, values ranging between 80% and 125% were within the WT range of activity, in which the trialanine mutation had *no effect* (green background). A mutant had a *mild effect* for values ranging between from >25 and <80 (orange background) and had a *strong effect* for values of 25% and below (red background).

See next page.

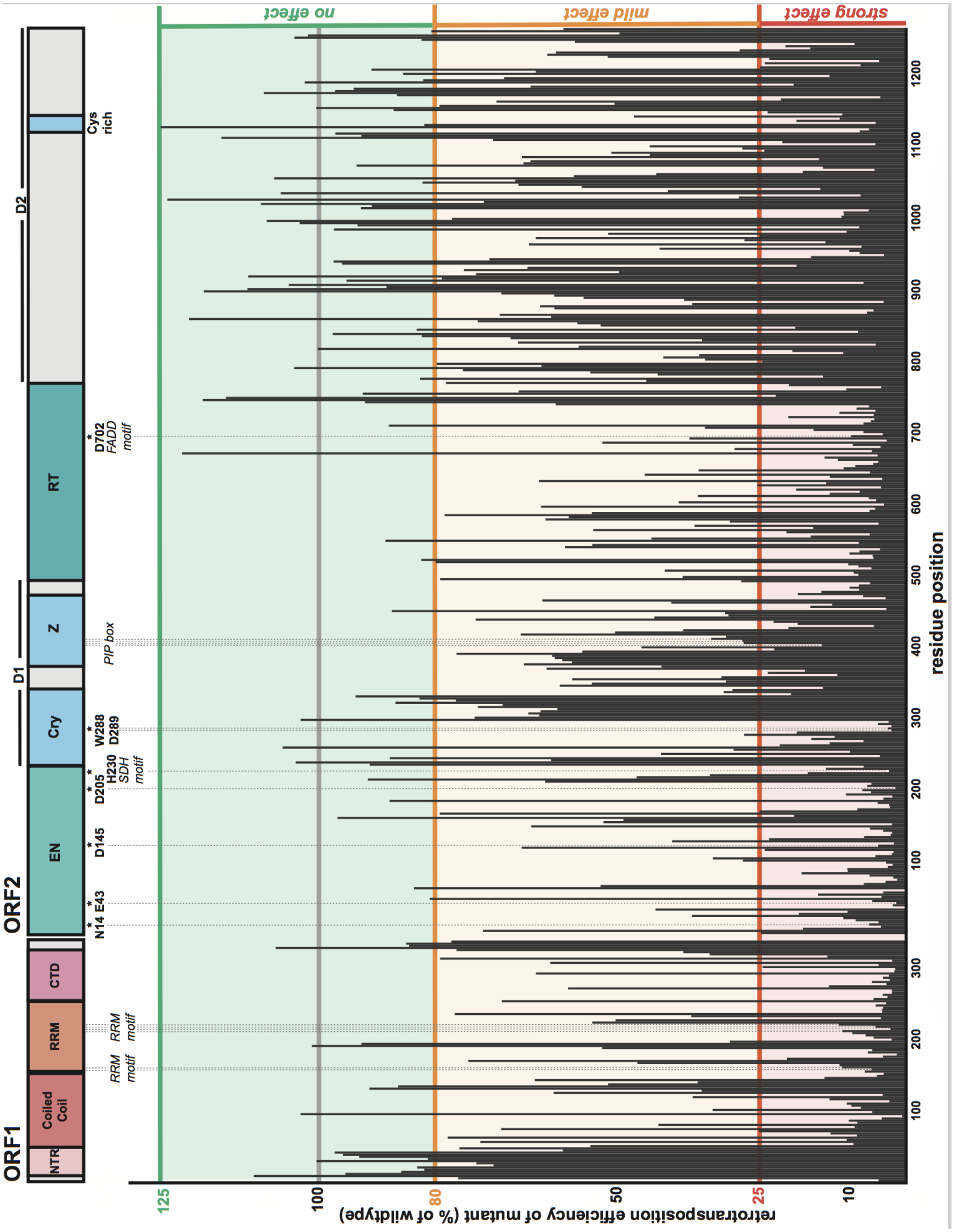
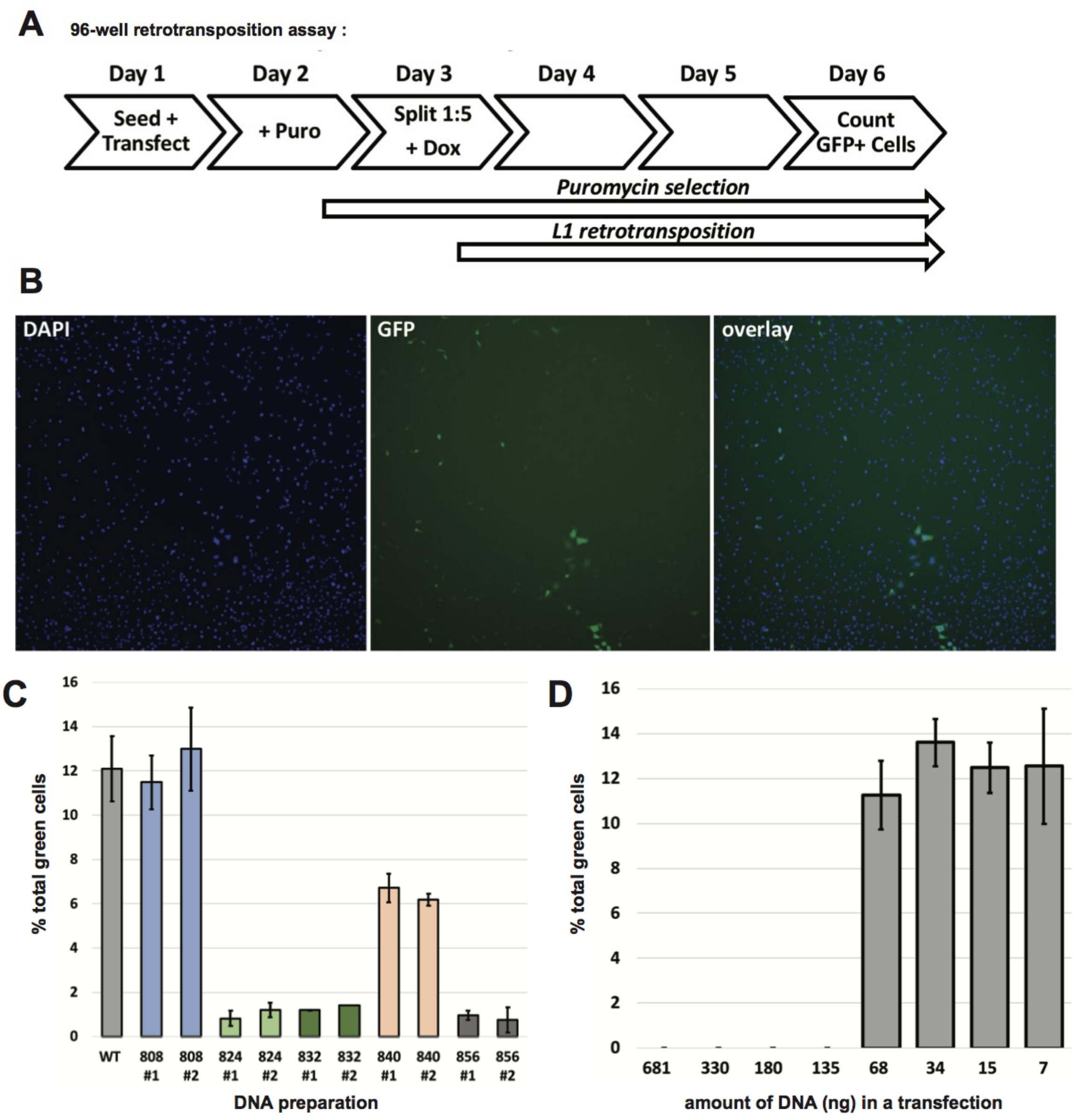

ORF1
ORF2
NTR
Coiled Coil
RRM
CTD
EN
Cry
Z
RT
D1
D2
Cys rich
*N14
*E43
*D145
*D205
*H230
SDH motif
*W288
D289
PIP box
*D702
FADD motif
RRM motif
RRM motif
no effect
mild effect
strong effect
125
100
80
50
25
10
retrotransposition efficiency of mutant (% of wildtype)
100
200
300
100
200
300
400
500
600
700
800
900
1000
1100
1200
residue position

Figure 3.7 : High-throughput microscopy-based measurement of retrotransposition.
(A) The protocol for the six-day 96-well, microscopy-based assay for plasmid transfection and measurement of the retrotransposition activity in human HeLa cells. Puromycin selection of the cells containing plasmid for five days was followed by dox-induced expression of L1 for four days. (B) Representative pictures of transfected WT constructs. This photo represents one quarter of one well in a 96-well plate. This shows two channels (DAPI and GFP) and the overlay. Blue (DAPI, alive) and green (GFP, retrotransposition-positive) cells were quantified. Raw wild-type retrotransposition values were reproducible and robust, usually showing ~13% GFP+ of live cells. (C) Because DNA was purified for 538 constructs in different batches, we wanted to be sure that there was no effect of the batch in which DNA was prepared on the retrotransposition efficiency. We measured retrotransposition for WT and two independent preparations of 5 different (color coded) mutants (chosen at random). (D) Comparison of transfecting different amounts of DNA. The fact that the retrotransposition frequency is independent of DNA concentrations within the range of concentrations used here shows that small fluctuations DNA stock concentration will not affect comparisons among mutants. Cells died when transfected with too much DNA and gave reproducible results over a concentration range of 7-68 ng (60ng was used for experiments unless indicated otherwise).

See next page.


A
96-well retrotransposition assay :
Day 1
Day 2
Day 3
Day 4
Day 5
Day 6
Seed + Transfect
+ Puro
Split 1:5 + Dox
Count GFP+ Cells
Puromycin selection
L1 retrotransposition
B
DAPI
GFP
overlay
C
% total green cells
16
14
12
10
8
6
4
2
0
WT
808 #1
808 #2
824 #1
824 #2
832 #1
832 #2
840 #1
840 #2
856 #1
856 #2
DNA preparation
D
% total green cells
16
14
12
10
8
6
4
2
0
681
330
180
135
68
34
15
7
amount of DNA (ng) in a transfection

% of 3xala mutations in each retrotransposition efficiency category:			
ORF1p			
	strong	***mild***	***none***
Full-length Protein	**53**	**31**	**16**
NTR	6	35	59
Coiled coil	56	35	9
RRM	70	24	6
CTD	55	32	13
ORF2p			
	strong	***mild***	***none***
Full-length Protein	**48**	**35**	**17**
EN	74	19	8
D1	38	47	15
Z	36	61	3
RT	62	28	10
D2	32	41	27

Table 3.4 : Impact on retrotransposition efficiency organized by protein domain.
The percentages of 3xAla variants that show a strong, mild, or no effect on retrotransposition efficiency are represented for both ORF1p and ORF2p. The values for the full-length protein and then for each domain are shown.

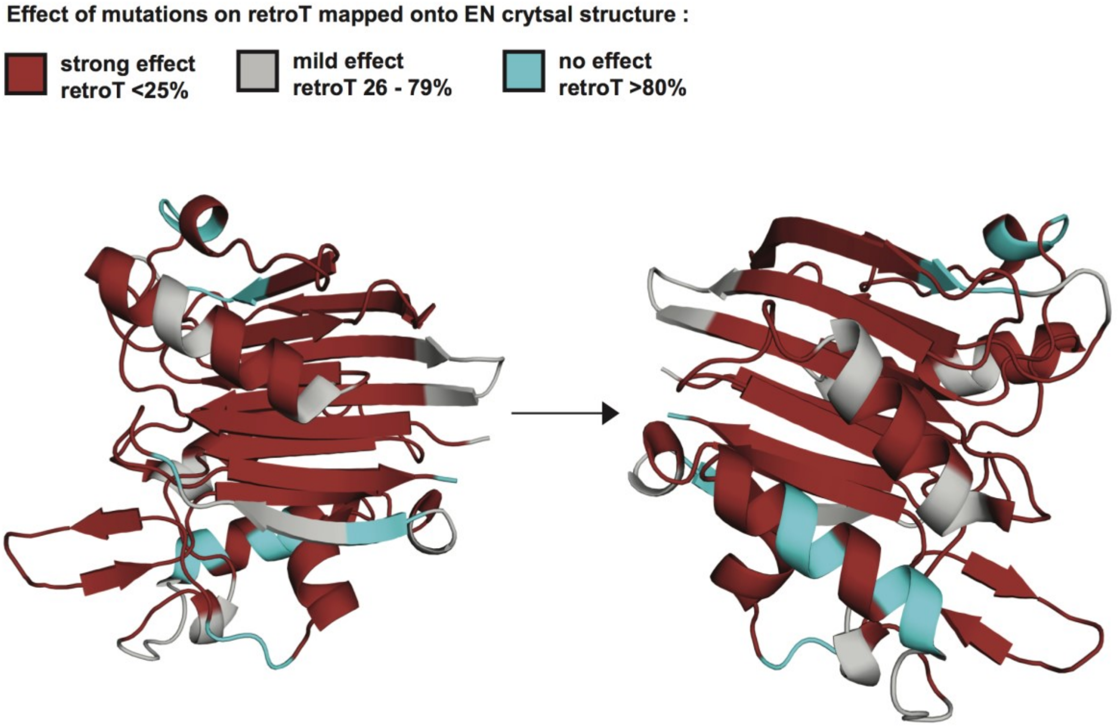

Effect of mutations on retroT mapped onto EN crytsal structure :
strong effect retroT <25%
mild effect retroT 26 - 79%
no effect retroT >80%

Figure 3.8 : Retrotransposition levels mapped onto EN 3D structure.
The crystal structure of the WT EN domain of ORF2 (PDB 1VYB) is shown and is color-coded to display retrotransposition efficiency of each trialanine variant. Red: strong impact on retrotransposition in red, Gray: mild impact, Cyan: no impact. (Note that this color scheme changes for the remainder of the chapter, as we focus on the ORF1p structures.)

Mobile elements that remain active in the human genome inspire consideration of the host-parasite arms race that has evolved [69]. L1 elements pose a strong risk to the host due to their strong mutagenic capacity. Conversely, the parasite benefits from having evolved to maintain a robust lifecycle. In this work, after studying the behavior of the trialanine mutant library of the L1-encoded ORF1 and ORF2 proteins, we found evidence for mechanisms though which both the host and the parasite benefit substantially. About half of the trialanine variants could not retrotranspose at the 25% the level of the WT element, with a substantial proportion with activity that fell to 5% or lower than that of WT, a very severe phenotype. Only 16% of the variants could retrotranspose at the WT level. For an L1 element to be so sensitive to mutation is extremely beneficial to the host.

o *Quantification of ORF1p variant protein abundance*

We quantified the relative levels of the ORF1p variants by Western blot analysis (Figure 3.9), using an antibody targeting a known in endogenous ORF1 ([14] targets amino acids 35 - 44 of human ORF1p; Milliprep, prod. number MABC1152). Due to substantial variations (2-fold) in ORF1p levels in replicate immunoblot experiments we treated the average protein abundance as binary, with a conservative cut off: high (above 25% that of WT) or low (below 25%). About 18% of the mutants (20/113) resulted in ORF1p reduction to <25% that of wildtype. All of these mutants with low ORF1p also showed complete loss of retrotransposition activity. Variants containing the epitope showed no signal when probed with this antibody. However, since they all showed WT of close-to WT levels of retrotransposition, we can confidently assume that their protein abundance levels were not in the low category. Of the variants that show low protein levels, all map to the RRM and CTD domains (12 and 8 mutants, respectively; Figure 3.10).

Interesting regions become most apparent when considering the visual representation of the data displayed on the full-length mORF1 structure. Figure 3.11 depicts how protein levels and retrotransposition activity trend along the wildtype crystal structure. WT retrotransposition levels were supported by only 16% (18/113) of the ORF1p mutants, which are those that tolerate alanine substitutions well. The proportion of which variant motifs fell into which category based on combined retrotransposition and protein abundance data are shown in Table 3.5.

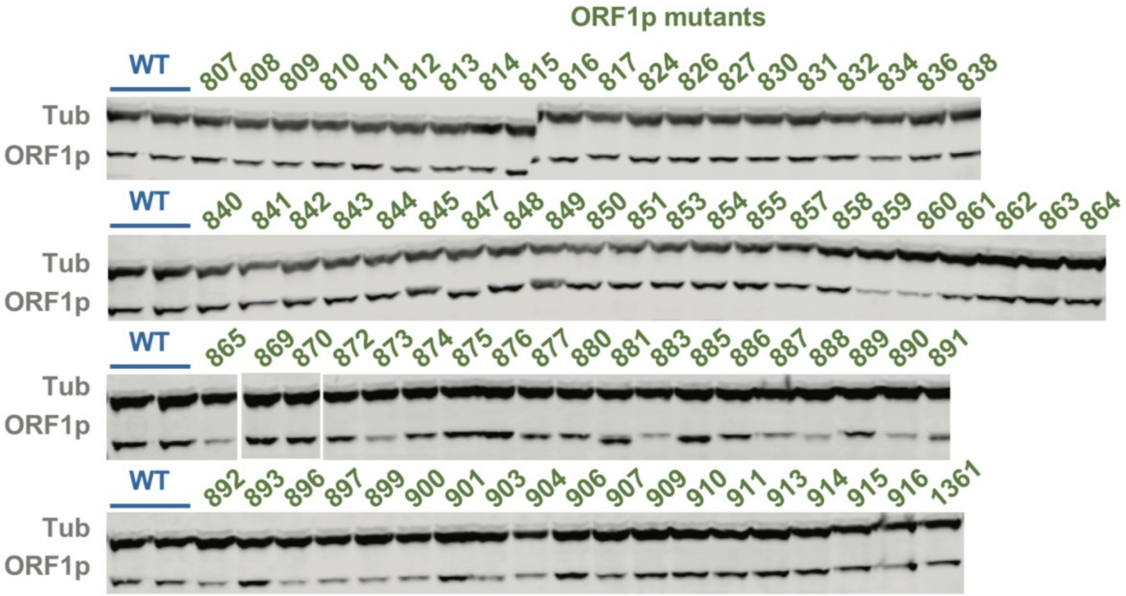


Figure 3.9 : Immunoblots of variants of ORF1p.

Representative immunoblots for WT pEA0264 and the ORF1p variants. Samples were prepared from 6-well plates of HeLa cell, the clarified lysate of which were probed with anti-ORF1 and anti-tubulin antibodies. HeLa cells lacking a plasmid reproducibly expressed ORF1p at a level of ~10% of pEA0264 (blot not shown).

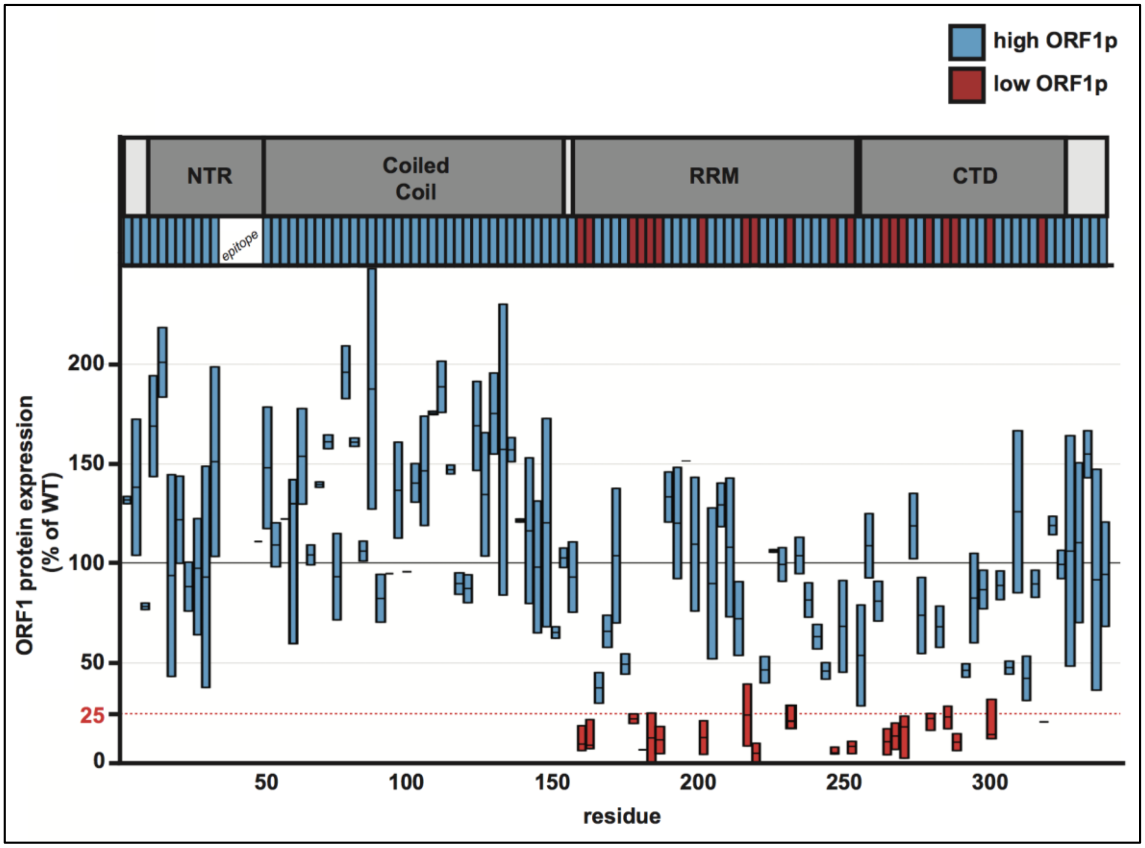


Figure 3.10 : Impact of ORF1 mutations on protein abundance.
The ORF1p schematic is shown at the top. Results from immunoblot analyses for each ORF1p variant are represented on the plot. Two measurements are shown for each variant, quantified from independent experiments. These values were background subtracted to remove signal corresponding to endogenous ORF1p expression. Protein levels are plotted on the Y axis and residue position is indicated on the x-axis. We observed some variability and thus plotted the range for each variant as a bar with a horizontal bar marking mean values. We refer to protein abundance in binary terms, as either high (+) or low (-), using 25% as the cutoff, indicated with a red dashed line. The variants are color coded in the bar below the ORF1p schematic to highlight which regions had high (blue) or low (red) protein levels.

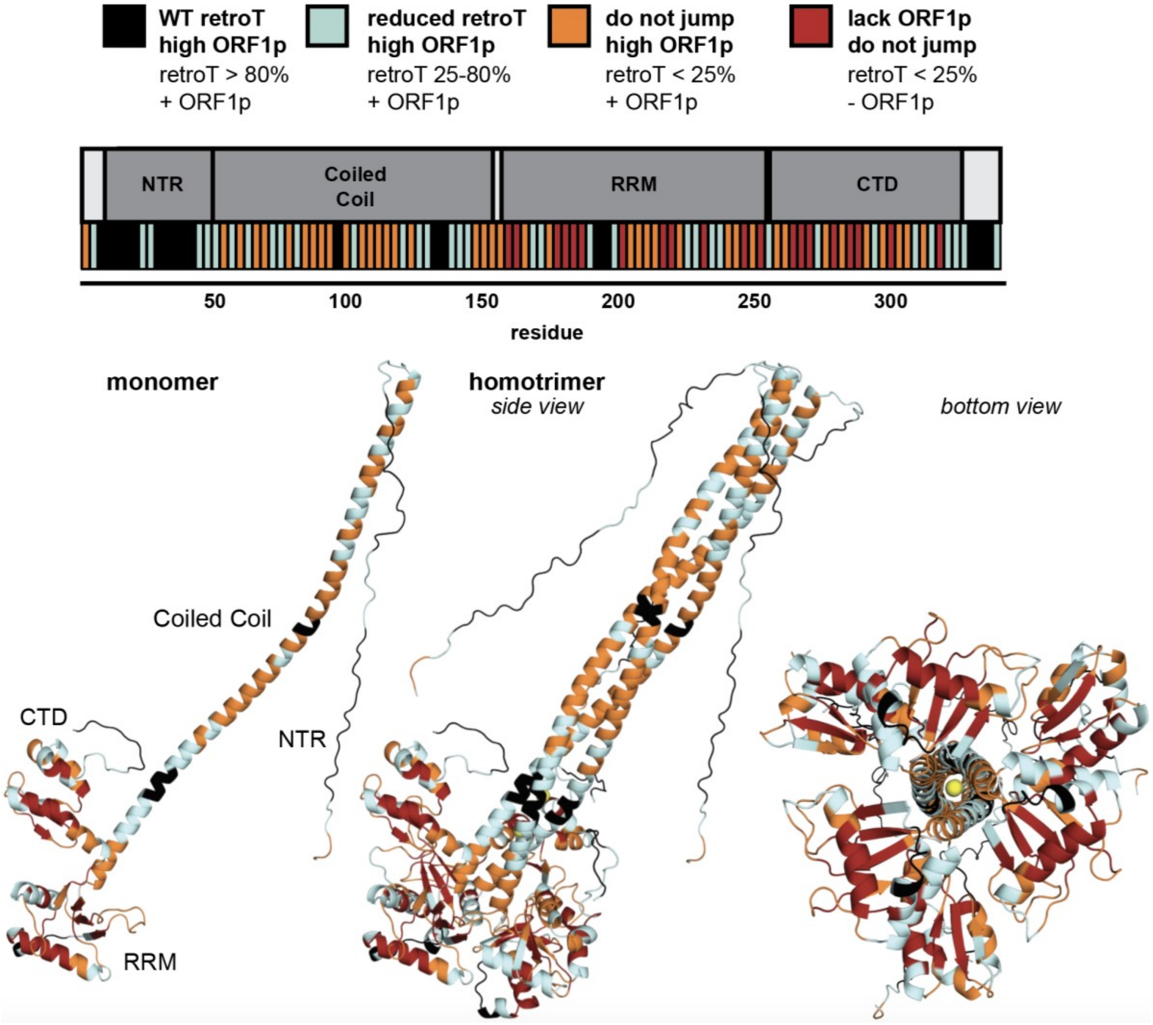


Figure 3.11 : Retrotransposition and protein abundance of mutants mapped onto ORF1p crystal structure.
The ORF1p mutants are divided into four categories and color coded, shown at the top. This provides a visual representation of retrotransposition efficiency and protein abundance, both along the linear schematic of ORF1 with the corresponding color coded bars as well as projected onto on the WT ORF1p monomer and trimer structures. The color code is as follows: high ORF1p and WT retrotransposition (black), high ORF1p and reduced retrotransposition (cyan), high ORF1p and no retrotransposition (orange), low ORF1p and no retrotransposition (red), chloride ions noted in the structure of Khazina and Weichenrieder (yellow), and the initial methionine (not mutated, white.)

Table 3.5 : Categories of ORF1p variants based on activity and protein abundance.
The percentages of ORF1p variants that show a strong, mild, or no effect on retrotransposition efficiency combined with whether or not they impact ORF1 protein abundance are represented in 4 categories (that correspond to Figure 3.11) for both ORF1p. The values for the full-length protein then for each domain are shown. (Note that the numbers for the CTD include the region through the C-terminus.

See next page.

% of 3xala mutations that landed in each final functional category:				
	ORF1p			
	WT retroT high ORF1p	**reduced retroT high ORF1p**	**no retroT high ORF1p**	**no retroT low ORF1p**
Full Length Protein	**16**	**31**	**35**	**18**
NTR	9	4	1	0
Coiled coil	3	12	18	0
RRM	2	7	9	11
CTD	3	8	8	7

- ***Probing which ORF1 mutants form RNPs with L1 RNA: pooled screening***

How ORF1p forms trimers, binds L1 RNA, and may form higher order structures makes examining the interaction of ORF1 within L1 RNP quite a complex system to study. As shown in Figure 3.12, we worked using the simplifying assumption based on prior work [30,31] that the *cis* model for ORF1p binding L1 RNA holds true and that each L1 RNP formed (i.e. one L1 RNA and its encoded proteins only are contained in each physical RNP particle) is physically independent of any another. We used this logic as the foundation for an experimental design testing the ability of each mutant to bind its own mutant L1 RNA as a measure of the capacity for L1 RNP formation, a critical step in the L1 lifecycle. This was our first step in probing the mechanism through which retrotransposition is reduced in some trialanine variants. This work will be presented in two parts: *Experiment 1)* This is a screen we did to test all 113 ORF1p mutants for RNP formation in a single extensive series of sequencing analyses, and *Experiment 2)* a variety of additional controls that we did to check the robustness of the dataset in its entirety. We are confident that this approach yields interesting insight into the biology of ORF1p motifs, but also consider some limitations.

The design for this experiment is simple : (1) overexpress the L1 plasmid(s) individually or in a pooled format of interest in human tissue culture cells, allow for ORF1p expression and binding to L1 RNA; (2) prepare a cell lysate; (3) extract total RNA from one fraction of lysate (in order to measure of total L1 RNA abundance), extract total plasmid DNA from one fraction of lysate (for normalization of RNA copy number to DNA copy number), and IP ORF1 from the remaining lysate; (4) discard the flow through and treat what remains on the beads for extraction of RNA. The latter two steps give a readout of the RNA associated with the pooled RNPs. The data are reported as the RNP formation capacity of an ORF1p variant compared to WT : IPd RNA / total RNA (normalized to total plasmid DNA).

When only WT L1 plasmid is transfected, WT ORF1p is IP'ed and the full L1 RNA is sequenced and assumed to have been in an RNP with that protein. The WT L1 RNA read depth represents the proportion of RNP complexes present in that sample. According to the cis-preference model, hypothetical Mut X and Mut Y plasmids **(**Figure 3.13**)** are cotransfected with the WT plasmid as part of the larger pool. Mut X is a mutant that binds RNA like wild-type whereas Mut Y does not. The ratios of the Mut X and Mut Y RNPs to WT RNP levels in the lysate can then be determined by comparing the number of RNA sequencing reads mapping to each of the three plasmids (because the 3xAla 9nt sequence is unique to each mutant). Under these assumptions, the pool composition should not affect these ratios. This is shown in Figure 3.13a (Figure 3.13b shows the details of the sequence analysis and is explained in the Methods).

Once we knew what to expect experimentally for the WT ORF1p alone, we completed *Experiment 1*, in which we interrogated each ORF1p variant using a pooled approach (Figure 3.13 and methods). The eight pools (M to T), each of which contained 12-14 mutant plasmids and the wild type pEA1348 plasmid, are described in Table 3.6 and are labeled M_{pool}-T_{pool}. For each variant, we measured total RNA abundance (Figure 3.14) and RNP formation ratios (Figure 3.15). Total RNA abundance data show that 90% of variants were well above 60% that of wildtype and the remaining 10% hovered around 60% that of wildtype. This dataset represents a direct measurement of the RNA levels for this subset of the mutants. Importantly, no mutants showed a strikingly depleted total RNA level, and thus an important conclusion of this experiment is that we see no indication that a lack of L1 mRNA (e.g. highly unstable mutant RNA) underlies any of the observed retrotransposition defects.

Thus, the phenotypes observed are likely due to direct effects on the mutated proteins - mutants that lead to problems with proper protein translation, folding, stability, or inter-macromolecular interactions. One might expect that ORF1p trimers would be most likely to form in the vicinity of translating L1 mRNA, and potentially bind to the non-translating part (e.g. the 3' ORF2 sequence), coating the RNA. It is possible that such binding might protect RNA from degradation. Any non-RNA binding ORF1p mutant would be unable to produce such structures. A skeptic of complete *cis-* action may also consider the following logic : if, in our system, there are *trans*-binding ORF1p variants, an ORF1p variant translated from one L1 RNA could protect the L1 RNA of a non-RNA binding ORF1p mutant. We would then need an "unpooled" one by one approach to test each variant's total RNA and plasmid DNA abundance to accomplish this, which due its labor intensiveness, has yet to be done.

Continuing with the results of *Experiment 1,* Figure 3.15 shows the RNP formation capacity of each variant, normalized to WT. As expected, no variant that produced exceptionally low amounts of protein was capable of forming RNPs. 12% of the ORF1p variants showed RNP formation at 30% or less that of WT. Combining all the data acquired to this point, the functional trends are shown in Figure 3.16. While most variants that are unable to form RNPs are located in the CTD and RRM domains, two variants stood out, mapping to the coiled coil domain, as unable to form RNPs. Interestingly, these included mutations of the five residues preceding the stammer and the first residue of the stammer motif. This supported the functional importance of the stammer region and indicated that perhaps the stammer played a key role in RNP formation. Also, overall, we concluded that the vast majority of ORF1p mutants form stable RNA and protein, bind L1 RNA, but still do not retrotranspose well. The question becomes, at the mechanistic level, why is this so?

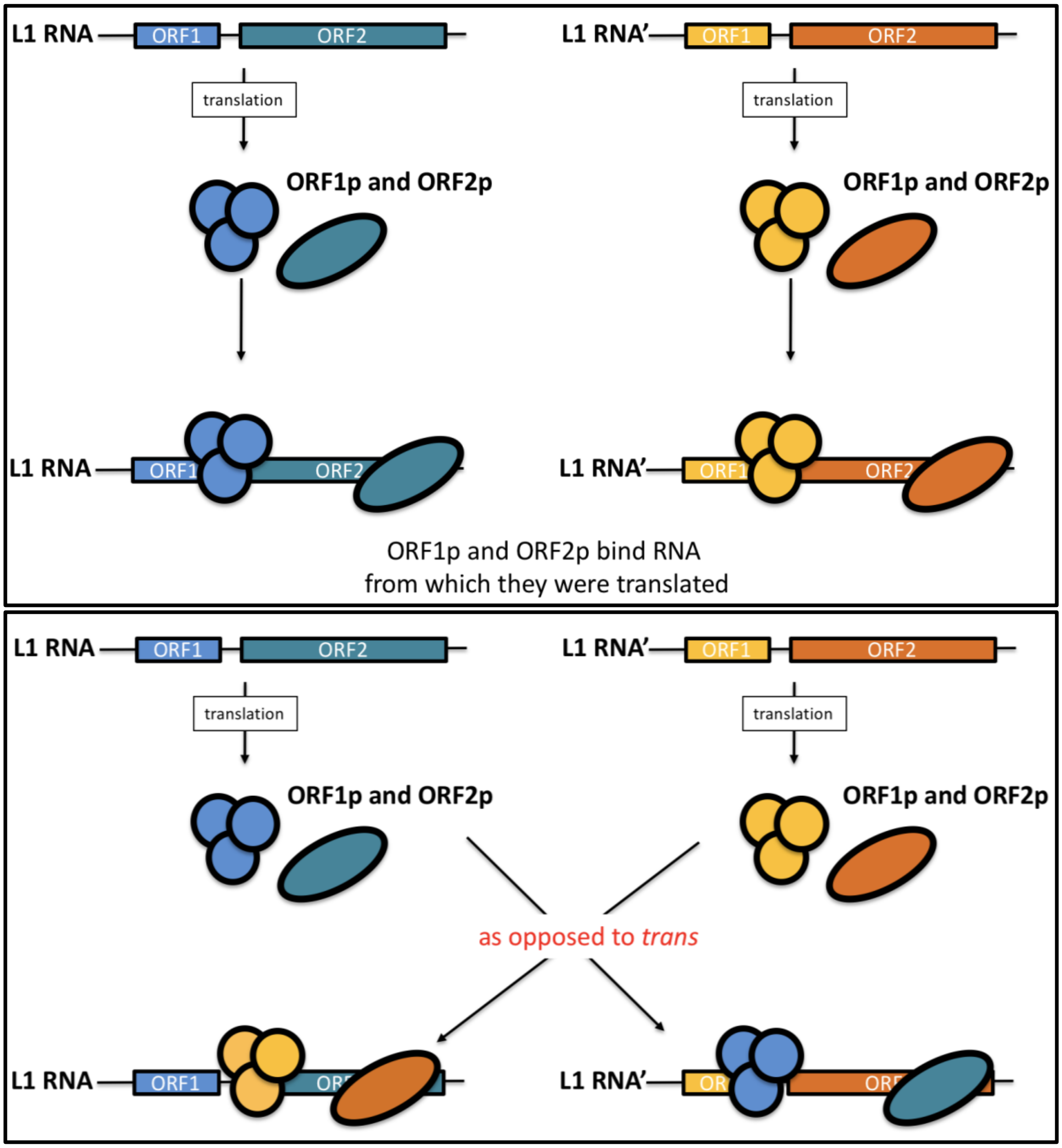
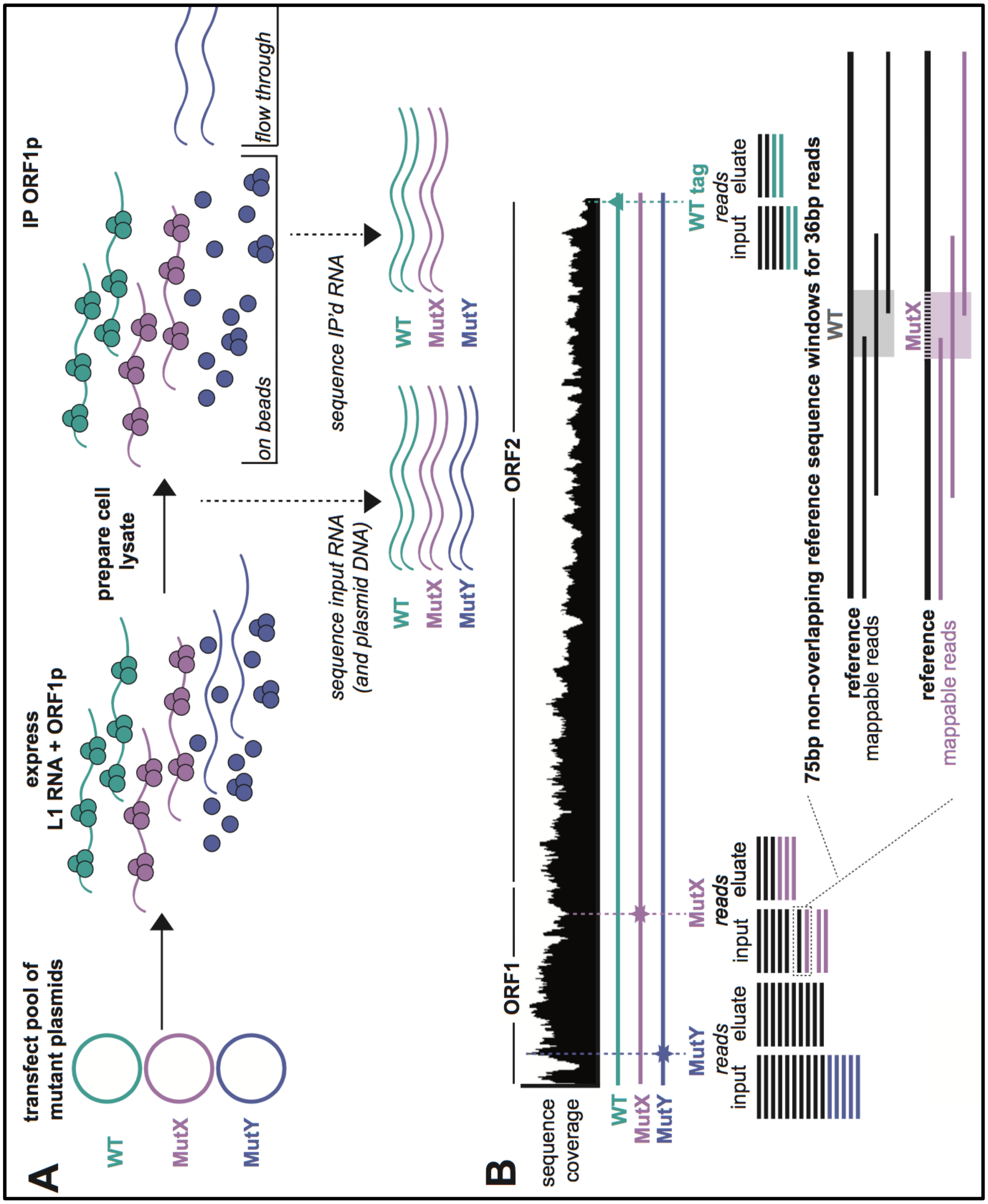
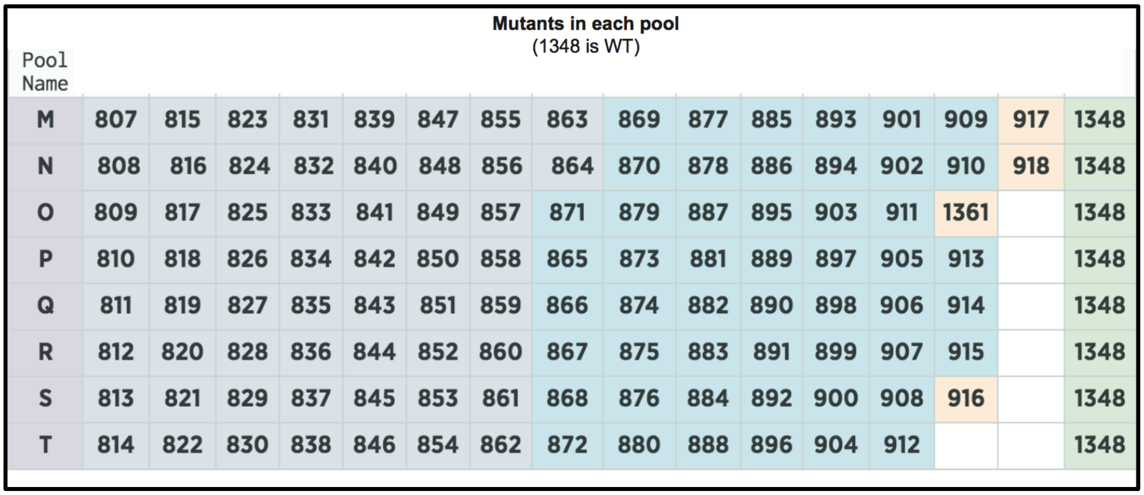


Figure 3.12: Schematic of the *cis*-preference model for binding of L1 RNA. There is strong evidence for and reasoning behind the *cis*- binding model of L1 proteins. The top shows schematically what cis-preference looks like (because this mechanism is important for the experiments done in this section), in which the L1 proteins bind the RNA from which they were translated. The bottom shows the opposite, namely *trans*-binding.

Figure 3.13 : Protocol for measuring an ORF1p variant's ability to form RNPs using pooled screening.
(A) Workflow for transfecting a pool of two mutants and the WT plasmid, expressing L1, and the IP is shown. In this depiction, MutX binds a WT level of RNA and MutY binds no RNA. The points at which samples are prepared for DNA and RNA sequencing are indicated. (B) Sequence coverage across L1 coding region is uneven; diagram depicts how reads were mapped and normalized.

See next page.


A
transfect pool of mutant plasmids
WT
MutX
MutY
express L1 RNA + ORF1p
prepare cell lysate
IP ORF1p
on beads
flow through
sequence input RNA (and plasmid DNA)
WT
MutX
MutY
sequence IP'd RNA
WT
MutX
MutY
B
sequence coverage
ORF1
ORF2
WT
MutX
MutY
MutY
reads
input
eluate
MutX
reads
input
eluate
WT tag
reads
input
eluate
75bp non-overlapping reference sequence windows for 36bp reads
reference
mappable reads
WT
reference
mappable reads
MutX



Pool Name	Mutants in each pool (1348 is WT)														
M	807	815	823	831	839	847	855	863	869	877	885	893	901	909	917	1348
N	808	816	824	832	840	848	856	864	870	878	886	894	902	910	918	1348
O	809	817	825	833	841	849	857	871	879	887	895	903	911	1361		1348
P	810	818	826	834	842	850	858	865	873	881	889	897	905	913		1348
Q	811	819	827	835	843	851	859	866	874	882	890	898	906	914		1348
R	812	820	828	836	844	852	860	867	875	883	891	899	907	915		1348
S	813	821	829	837	845	853	861	868	876	884	892	900	908	916		1348
T	814	822	830	838	846	854	862	872	880	888	896	904	912			1348

Table 3.6 : Composition of mutant pools for RNP studies (*Experiment 1*). The composition of each of the pools (M_{pool}-T_{pool}), each mutant is listed using the pEA clone number as the ID. This experiment was done fully for the complete series of variants of ORF1p (DNA, input RNA, and IP'd RNA sequencing) once.

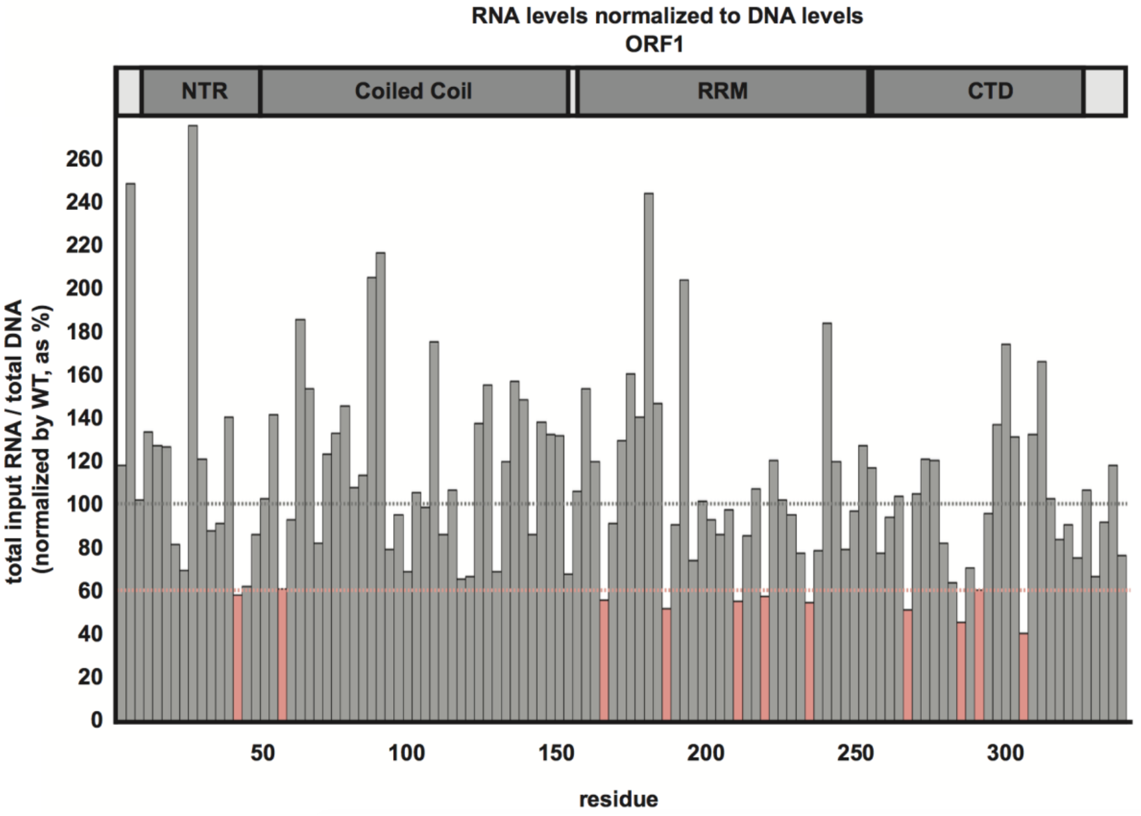


Figure 3.14 : RNP formation of each ORF1p variant for RNP studies (*Experiment 1*).

The RNP formation of each ORF1p variant is shown as % of WT. The fraction of mutant L1 RNA pulled down in the IP, normalized by WT is shown. The gray dashed line indicates 100%, WT RNA bound, light red dashed line indicates 60% (light red bars are variants with values between 30 and 60%), and the dark red dashed line indicates 30% (dark red bars are variants with values between 30 and 60%).

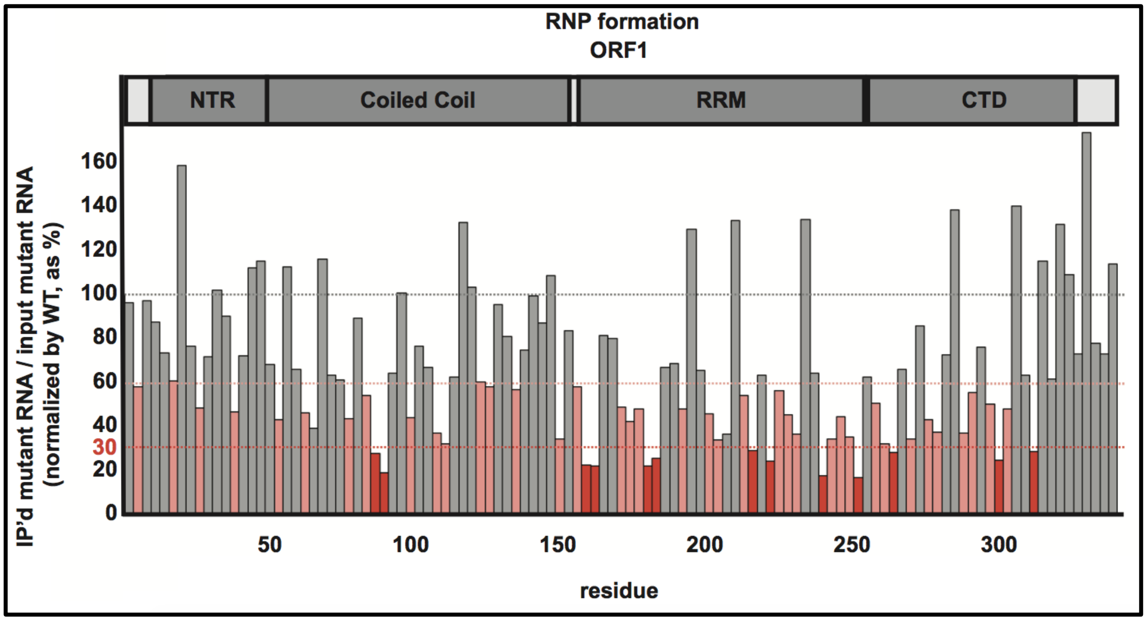


Figure 3.15 : RNP formation of each ORF1p variant for RNP studies (*Experiment 1*).

The RNP formation of each ORF1p variant is shown as % of WT. The fraction of mutant L1 RNA pulled down in the IP, normalized by WT is shown. The gray dashed line indicates 100%, WT RNA bound, light red dashed line indicates 60% (light red bars are variants with values between 30 and 60%), and the dark red dashed line indicates 30% (dark red bars are variants with values between 30 and 60%). There is no endogenous ORF1p background subtraction.

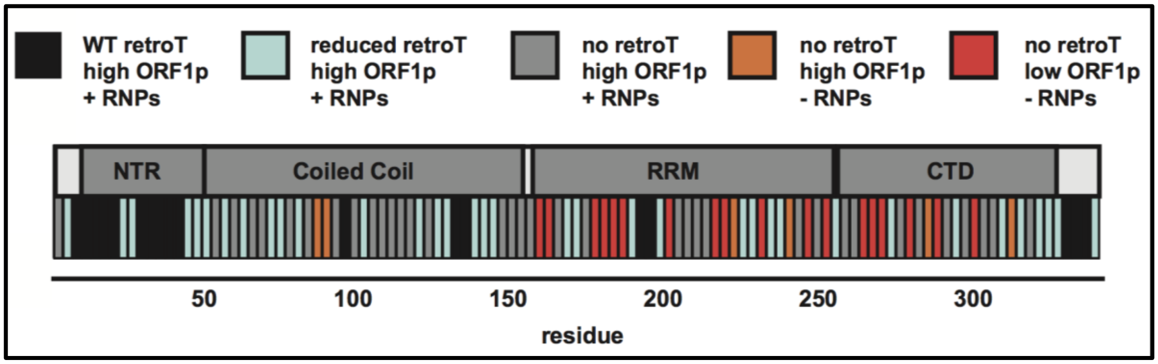


Figure 3.16 : Functional categories of ORF1p mutants based on RNP studies (*Experiment 1*).

The ORF1p mutants are divided into six categories and color-coded. This creates a visual representation of retrotransposition efficiency, protein abundance, and L1 RNA binding, along the linear map of ORF1p with the corresponding color-coded bars. *Retrotransposition categories* : WT retrotransposition (>80%), reduced retrotransposition (>25%-<80%), do not retrotranspose (<25%). *Protein abundance categories* : high ORF1p (>30%), low ORF1p (<30%). *RNP formation categories* : + RNPs (forms RNPs >30%), - RNPs (forms RNPs <30%) *The color code is as follows:* WT retrotransposition, high ORF1p, + RNPs (black), reduced retrotransposition, high ORF1p, + RNPs (light cyan), do not retrotranspose, high ORF1p, + RNPs (gray), do not retrotranspose, high ORF1p, -RNPs (pink), do not retrotranspose, low ORF1p, + RNPs (orange), and do not retrotranspose, low ORF1p, - RNPs (red).

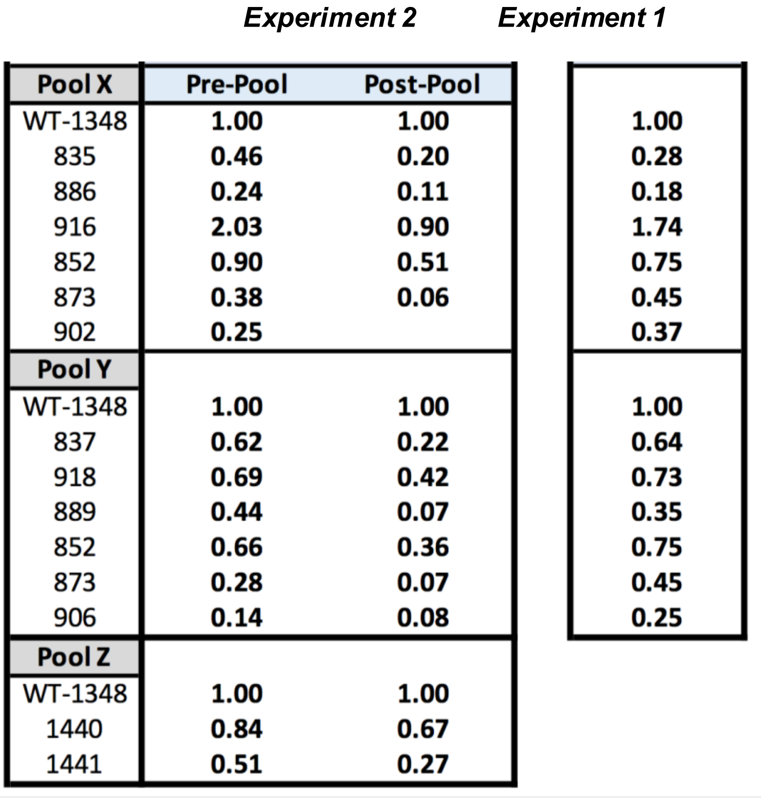
Continuing into *Experiment 2* of the ORF1p RNP formation analyses, we carried out some very important controls. We made new samples and went through the previously described experimental process, but using a series of pools designated X_{pool} and Y_{pool}. Both sets contained the WT plasmid, two of the same variants (to show that the RNP levels that we saw were reproducible, independent of pool content), and also a few other variants. This time, for each X_{pool} and Y_{pool}, we compared the RNP formation values from two types of sample preparations. The "Pre-Pool" sample had all plasmids in a given pool combined and transfected into cells together, such that all were expressed in one batch of cells (as was done in *Experiment 1).* The "Post-Pool" sample had each plasmid in that pool transfected separately and the pool was not combined until the step of lysate preparation. Thus, only one L1 trialanine variant plasmid was expressed through the formation of RNPs in each cells. In keeping with the *cis-* model, we anticipated that the "Pre-Pool" and "Post-Pool" samples should result in the same RNP formation values. (For technical reasons, we assumed that the ratio of total RNA (not measured in this experiment) ratio to total DNA (measured) was the same as measured in *Experiment 1*.)

Finally, the last pool, Z_{pool}, also underwent the "Pre-Pool" and "Post-Pool" preparations. This pool contained the WT plasmid and two specially made constructs, one that lacked (pEA1440) and one that fully mutated (pEA1441) (refer to Table 3.2) the epitope (which lies at the N-term in the NTR domain) of the anti-ORF1 antibody being used in the IP. Both 1440 and 1441 retrotranspose, express protein and should bind L1 RNA. If all these experiments are proceeding as hoped, we expect that these variants would not bind the antibody and thus no corresponding L1 RNA should be detected.

The RNP formation results are shown in Figure 3.17. Generally speaking, the trends in RNP formation are similar comparing the results obtained in *Experiment 1* and the *Experiment*

2 "Pre-Pool" and "Post-Pool" samples, only the observed phenotype is often quite a bit stronger in the "Post-Pool" subset. This result is consistent with cis action being a tendency but not an absolute phenomenon.

The Z_{pool} result revealed the biggest caveat in interpreting these data. A substantial amount of L1 RNA corresponding to plasmids that lacked the epitope recognized by the anti-ORF1p antibody used for the IP were recovered in all the samples. This could mean a few things. First, it indicates that the assay is not as clean as we need it to be. Second, it might indicate that WT ORF1p is able to bind the mutant RNAs of pEA1440 and 1441 substantially in *trans*, which would lead to "subunit mixing" (i.e. each RNP consists of a mix of WT and mutant ORF1p subunits) severely limiting our ability to interpret these data. The last potential explanation centers around the potential of ORF1p to form higher order meshworks (i.e. inter RNP interactions as shown in Figure 3.2). If these configurations are forming ad can be immunoprecipitated, each of the ORF1p variant trimers might have been binding purely in *cis* to their L1 transcript, but the "open" conformations of WT, 1440, 1441, might allow for formation of higher order linear and meshwork structures to form between them. Thus, perhaps they were in fact binding their own RNA in *cis,* but in this system, IP of the epitope-containing WT construct was enough to bring down the whole heterogeneous protein meshwork. No matter what is truly happening, more single and various combinations of multiple mutants will need to be analyzed to get convincing answers in terms of using these mutants to query cis action. This will also help avoid the potential caveat that certain mutations could create false negatives if the trialanine mutation removes a motif essential for *cis*-binding.

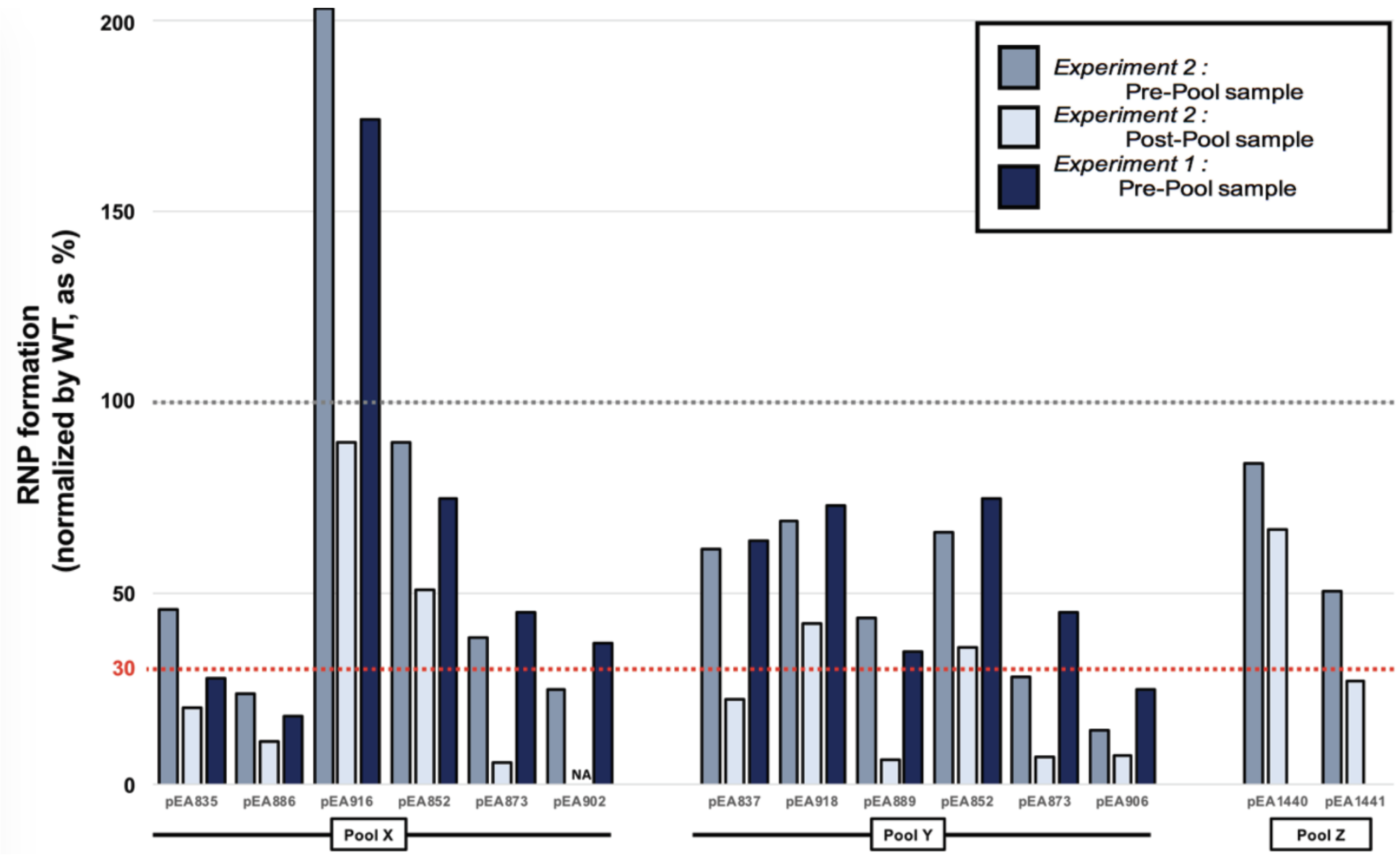


	Experiment 2		Experiment 1
Pool X	**Pre-Pool**	**Post-Pool**	
WT-1348	**1.00**	**1.00**	**1.00**
835	**0.46**	**0.20**	**0.28**
886	**0.24**	**0.11**	**0.18**
916	**2.03**	**0.90**	**1.74**
852	**0.90**	**0.51**	**0.75**
873	**0.38**	**0.06**	**0.45**
902	**0.25**		**0.37**
Pool Y			
WT-1348	**1.00**	**1.00**	**1.00**
837	**0.62**	**0.22**	**0.64**
918	**0.69**	**0.42**	**0.73**
889	**0.44**	**0.07**	**0.35**
852	**0.66**	**0.36**	**0.75**
873	**0.28**	**0.07**	**0.45**
906	**0.14**	**0.08**	**0.25**
Pool Z			
WT-1348	**1.00**	**1.00**	
1440	**0.84**	**0.67**	
1441	**0.51**	**0.27**	

Table 3.7 : RNP formation of each ORF1p variant for RNP studies (*Experiment 2*).
The raw data for controls done in addition to *Experiment 1* is shown. The left column displays pool names and the pEA IDs of the constructs used in those pools. The middle column displays the data for RNP formation, normalized to DNA (total RNA levels) and to WT, indicated at 1 (100% RNP formation). DNA levels were quantified for these samples, but total RNA was not. Therefore the relative amount of RNA to DNA that was measured in Figure 3.12 (from *Experiment 1*) were used for normalization here. The right column shows RNP formation found for the given constructs when they were measured in *Experiment 1* (only Pre-Pool samples) pools M_{pool}-T_{pool}.

Figure 3.17 : RNP formation of each ORF1p variant for RNP studies (*Experiment 2*).
This shows the data in Table 3.5 shown graphically. The pools (X_{pool}-Z_{pool}) are indicated along the bottom (as well as the individual constructs in each pool). The gray dotted line indicates the WT level of RNA recovered in the RNP fraction and the red dotted line refers to the cutoff used for analysis of the results from *Experiment 1*. The legend indicates the color code used for each experimental condition.

See next page.


RNP formation
(normalized by WT, as %)
200
150
100
50
30
0
Experiment 2 : Pre-Pool sample
Experiment 2 : Post-Pool sample
Experiment 1 : Pre-Pool sample
pEA835
pEA886
pEA916
pEA852
pEA873
pEA902
NA
pEA837
pEA918
pEA889
pEA852
pEA873
pEA906
pEA1440
pEA1441
Pool X
Pool Y
Pool Z

- ***Investigation of trends in nucleolar localization of certain ORF1p trialanine mutants***

We used IF to probe the localization of a subset of the ORF1p trialanine variants that span the entire protein and also fit various functional categories, as defined so far. WT ORF1p, which is most abundantly found in the cytoplasm, was compared to 42 of the 113 ORF1p variants. While WT ORF1p and most of the variants did not localize to the nucleolus, we did see two main patches of motifs in ORF1p in which nucleolar localization is distinctively present for those variants in a subset of cells and experiments; when seen however this localization was extremely striking (Table 3.8 and Figure 3.18). The first patch is in the coiled coil domain and, intriguingly, overlaps with variants that contain the stammer region of the coiled coil. Sixteen other variants in the coiled coil were tested and show no phenotype. Ongoing work (not shown) indicates a strong nucleolar phenotype for the variant preceding and the two variants containing the structurally-critical stammer residues (pEA0835-837). Interestingly, these particular variants (pEA835 and 836) are the only variants in the coiled coil that could not form proper RNPs over 30% the levels of WT. We have built and are testing two constructs that have the three-residue stammer (only) deleted as well as mutated to a trialanine sequence. We know that making both of these adjustments kills retrotransposition [62]**.** The second patch that contains the nucleolar phenotype lies within the CTD, which is critical for RNA-binding.

The localization of a subset of ORF1p variants to the nucleolus is likely the result of their altered binding affinities for nucleic acid or protein partners. *Alu*, a SINE, has been shown to accumulate in the nucleolus where it plays a role in nucleolar size, activity, and localization [70]. Additionally, structural work has indicated that the *Alu* RNP complex with SRP9 and

SRP14 is able to stall translational elongation in a conformation that may allow *Alu* RNA to interrupt the cis-preference of L1 proteins and co-opt ORF2p [71]. Perhaps the various trialanine mutations in ORF1p alter trimer assembly and loading onto L1 RNA and thus allow more mutant ORF1p to bind to *Alu* RNAs and localize to the nucleolus. Notably ORF1p has been shown not to be required for *Alu* retrotransposition (Dewannieux et al. 2003); however, ORF1p supplementation was shown to enhance *Alu* retrotransposition (Wallace et al. 2008). Thus mutations in ORF1p could modulate *Alu* retrotransposition efficiency through either co-translational RNA-binding changes or changes in RNP shuttling. In fact, this is a part of the inspiration for what would be one of the most relevant and exciting applications of this library: since L1 mobilizes *Alu*, if we could engineer a robust fluorescent reporter for *Alu* insertion efficiency, we could test how the library of the L1 trialanine variants impacts Alu activity in a very similar way to how we measured L1 retrotransposition. We could map regions critical for *Alu*-binding and function.

When considering redistribution from a punctate nuclear pattern to diffuse nucleolar localization, the similarities to the PML protein are also compelling. The PML gene is a tumor suppressor that is involved in the pathogenesis of acute promyelocytic leukemia (APL). The PML protein localizes to multi-protein sub-nuclear structures known as PML-Nuclear Bodies (PML-NBs) that have been shown to be important in p53 stabilization and activation [72]. In response to a number of cellular stresses such as DNA damage and transcriptional inhibition, the PML protein has been shown to change its nuclear localization, including nucleolar relocalization in some cases [73]. Notably, PML-NBs have also been implicated in the regulation of reverse transcription of HIV-1 and endogenous retroviruses [74].Nucleolar localization of mutant ORF1p may then be a result of decreased binding affinity for PML-NB proteins, which would then allow ORF1p to freely redistribute in the nucleus. Alternatively, specific ORF1p mutations could create or mimic cellular stress conditions that would promote PML redistribution to the nucleolus, which could in turn drive nucleolar relocalization of ORF1p.

Another candidate protein interaction that could be at play is that between ORF1p and Rad18. Rad18 is a DNA damage-responsive E3 ubiquitin ligase that complexes with Rad6 and monoubiquitinates PCNA to promote DNA polymerase switching and downstream replication at DNA lesions [75]. Notably Rad18 was recently shown to restrict L1 and *Alu* retrotransposition through direct binding to ORF1p [76]. Like a number of other DNA damage response pathway proteins, Rad18 has been shown to localize to the nucleolus in a cell cycle-dependent manner [77]. Hence, mutations in ORF1p could modulate the binding affinity of Rad18 for ORF1p, which could promote strong nucleolar localization of ORF1p potentially with cell cycle dependence.

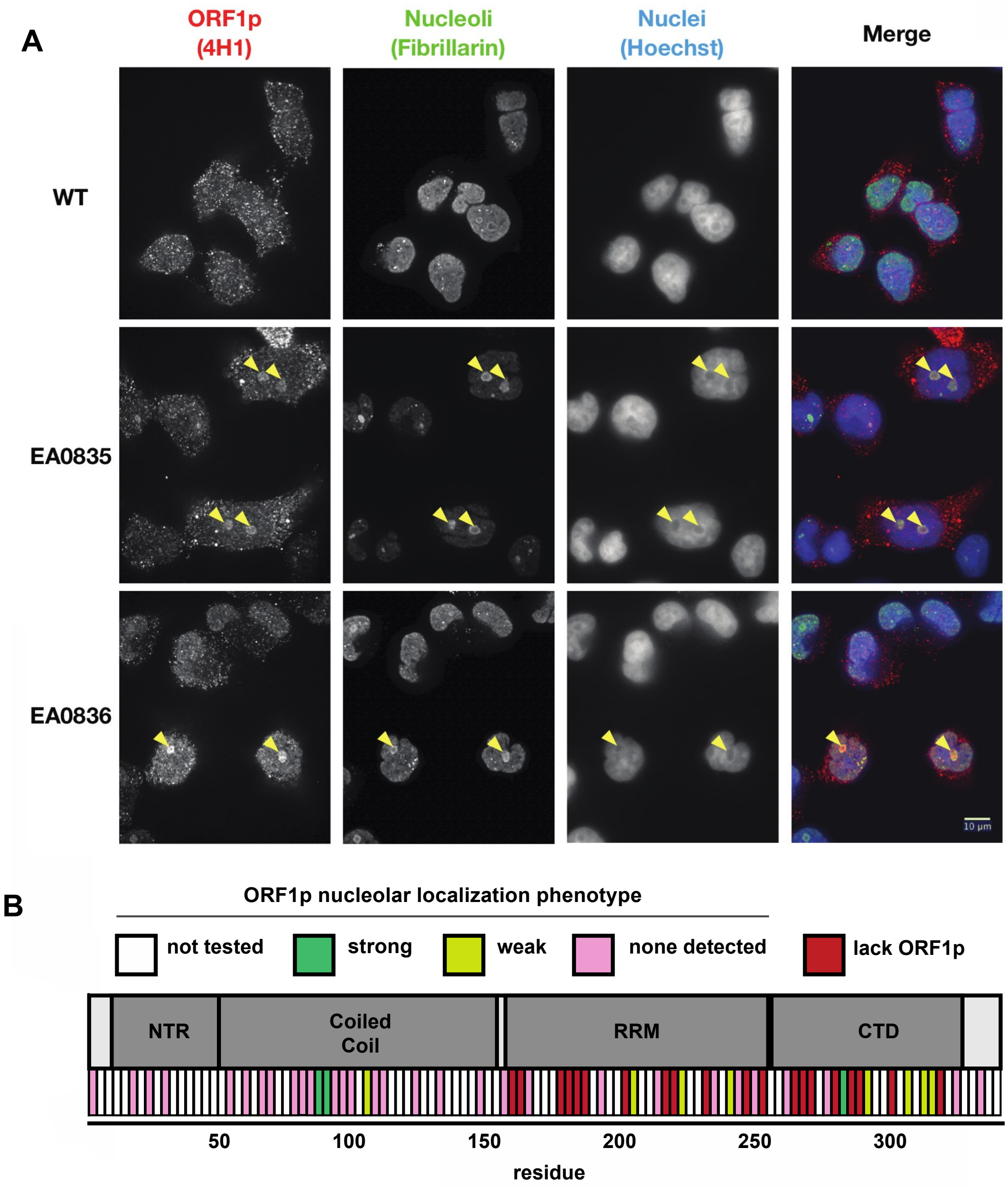
Table 3.8 : An analysis of ORF1p variants that localize to the nucleolus. This is a comparison of nucleolar localization phenotype by immunocytochemistry of ORF1p. We tested 42 of the 113 ORF1p mutants to study cellular localization of ORF1 protein. Column one shows the construct tested, columns two indicates the domain in which it lies, column three shows the retrotransposition efficiency (++ : >80% of WT, + : 25-80% of WT, and – : <25% of WT), column four indicates the number of times that variant was tested (in independent transfections), and column five shows, on a binary scale, whether we determined there to be any nucleolar localization. Each variant expresses ORF1protein (above the 25% of WT threshold) except for that which is indicated with ** in column one. The notes in the last two columns describe the strongest phenotype seen, first in terms of approximate percentage of cells that exhibited the phenotype and, second, in terms of the qualitative brightness of the nucleolar ORF1p signal compared to WT. Experiments performed by Srinjoy Sil.

See next page.

	ORF1p domain	retroT category	# of measurements	nuclear phenotype	strongest phenotype observed	
					% nucleolar ORF1 positive cells	positive cell nucleolar ORF1p brightness
WT			4	-		
pEA807_2_4_GKK	**NTR**	-	1	-		
pEA812_17_19_ASP		++	1	-		
pEA814_23_25_ERS		+	1	-		
pEA816_29_31_ATE		++	1	-		
pEA824_53_55_SEL	**CC**	-	1	-		
pEA826_59_61_IQT		-	1	-		
pEA828_65_67_EVE		-	1	-		
pEA829_68_70_NFE		-	1	-		
pEA832_77_79_ITR		-	1	-		
pEA833_80_82_ITN		+	1	-		
pEA834_83_85_TEK		-	1	-		
pEA835_86_88_CLK		-	2	+	10	bright
pEA836_89_91_ELM		-	4	+	20	bright
pEA837_92_94_ELK		-	3	-		
pEA838_95_97_TKA		++	1	-		
pEA839_98_100_REL		-	1	-		
pEA841_104_106_CRS		-	1	+	< 1	dim
pEA842_107_109_LRS		-	1	-		
pEA843_110_112_RCD		-	1	-		
pEA847_122_124_EDE		-	1	-		
pEA850_131_133_EGK		++	1	-		
pEA855_146_148_LQE		-	1	-		
pEA858_155_157_RPN		-	1	-		
pEA861_164_166_PES	**RRM**	-	1	-		
pEA866_179_181_QDI **		-	2	-		
pEA870_191_193_RQA		++	1	-		
pEA874_203_205_TPQ		-	1	+	< 1	dim
pEA877_212_214_ATP		-	1	-		
pEA880_221_223_FTK		-	2	+	< 1	bright
pEA884_233_235_AAR		+	1	-		
pEA886_239_241_RVT		-	2	+	< 1	bright
pEA887_242_244_LKG		-	1	-		
pEA889_248_250_RLT		-	1	-		
pEA893_260_262_ARR	**CTD**	-	1	-		
pEA898_275_277_NFQ		-	1	-		
pEA900_281_283_SYP		-	2	+	20	bright
pEA903_290_292_SEG		-	1	+	1	dim
pEA908_305_307_DFV		-	1	+	1	bright
pEA910_311_313_PAL		-	2	+	1	bright
pEA911_314_316_KEL		+	1	+	< 1	dim
pEA914_323_325_MER		+	1	-		
pEA917_332_334_LQN		++	1	-		

Figure 3.18 : IF analysis reveals intriguing nucleolar localization of some ORF1p variants.
Certain ORF1p mutations can cause localization to the nucleolus. (A) Representative images of immunostained fixed Hela cells expressing wild-type (WT) L1 or one of two L1 ORF1p mutants (EA0835 & EA0836), which exhibit the "strong" nucleolar localization phenotype (Table 3.8). Antibody target names are indicated along the top (fibrillarin marks the nucleolus and Hoechst marks nuclear DNA) and the corresponding pictures are colored according to the colors used in the merged pictures. Scale bar = 10 μm. Yellow arrowheads indicate nucleoli with diffuse ORF1p localization. (B) The ORF1 schematic shows which variants show which nucleolar phenotype: variant not tested (white), strong (seen bright and >10% of cells; dark green), weak (light green), none detected (pink), and those that lack ORF1p (protein <25% that of WT; red). Experiments by Srinjoy Sil.

See next page.


A
ORF1p
(4H1)
Nucleoli
(Fibrillarin)
Nuclei
(Hoechst)
Merge
WT
EA0835
EA0836
10 μm
B
ORF1p nucleolar localization phenotype
not tested
strong
weak
none detected
lack ORF1p
NTR
Coiled
Coil
RRM
CTD
50
100
150
200
250
300
residue

- ***Trends in amino acid conservation and sensitivity to mutation across ORF2p***

To gain a perspective on amino acid conservation in ORF2p, we aligned the human ORF2 protein sequence to 14 diverse mammalian sequences as well as others from more distant vertebrates (12 frog, 17 zebrafish, and 12 lizard sequences; Tables 3.9 and Table 3.10). Until now, conservation of functional residues has been integral to identifying regions of ORF2p indispensable for L1 activity. For the regions of unknown function, we reasoned that retrotransposition activity observed across ORF2p would help guide future studies.

Figure 3.19 shows the conservation and retrotransposition frequencies of each ORF2p variant. This highlights the variants that showed high sensitivity to mutation but lack sequence conservation and draws the eye to what we call the "Star Cluster", contained in the window of residues F952 – C1020, a region with a high density of motifs with this phenotype.

MAMMALS	VERTEBRATES		
>HUMAN_L1RP	>FROG_L1-6/1-1248	>ZEBRAFISH_L1-1A/1-1278	>LIZARD_L1_AC1/1-1256
>RABBIT/1-1274	>FROG_L1-32/1-1260	>ZEBRAFISH_L1-7B/1-1270	>LIZARD_L1_AC3b/1-1261
>PIG/1-1272	>FROG_L1-17/1-1244	>ZEBRAFISH/L1-7C_corr/1-1271	>LIZARD_L1_AC7/1-1250
>COW/1-1272	>FROG_L1-11/1-1243	>ZEBRAFISH_L1-13A/1-1270	>LIZARD_L1_AC8/1-1250
>DOG/1-1275	>FROG_L1-35/1-1248	>ZEBRAFISH_L1-13D/1-1268	>LIZARD_L1_AC9/1-1250
>PANDA/1-1275	>FROG_L1-18/1-1254	>ZEBRAFISH_L1-13C/1-1271	>LIZARD_L1_AC11/1-1256
>HORSE/1-1293	>FROG_L1-15/1-1244	>ZEBRAFISH_L1-13B/1-1264	>LIZARD_L1_AC12/1-1270
>LEMUR/1-1291	>FROG_L1-29/1-1247	>ZEBRAFISH_L1-1B/1-1264	>LIZARD_L1_AC14/1-1251
>ELEPHANT/1-1271	>FROG_L1-38/1-1257	>ZEBRAFISH_L1-16B/1-1266	>LIZARD_L1_AC15/1-1243
>HYRAX/1-1274	>FROG_L1-46/1-1284	>ZEBRAFISH_L1-1D/1-1259	>LIZARD_L1_AC17/1-1251
>ARMADILLO/1-1272	>FROG_L1-47/1-1266	>ZEBRAFISH_L1-10B/1-1272	>LIZARD_L1_AC18/1-1240
>MOUSE_A/1-1281	>FROG_L1-39/1-1263	>ZEBRAFISH_L1-11Aa/1-1244	>LIZARD_L1_AC20/1-1243
>RAT/1-1286		>ZEBRAFISH_L1-8/1-1260	
>OPOSSUM/1-1268		>ZEBRAFISH_L1-12A/1-1261	
		>ZEBRAFISH_L1-12B/1-1271	
		>ZEBRAFISH_L1-6/1-1255	
		>ZEBRAFISH_L1-17B/1-1262	

Table 3.9 : IDs of fifty-five LINE-1 ORF2p sequences used for phylogenetic conservation analysis.

These sequence IDs, shown in the table above, were provided by Stephane Boissinot and Oliver Weichenrieder. These were carefully curated as the most reliable ORF2p sequences. These were used for data presented in Table 3.10 and Figure 3.18.

Table 3.10 : Phylogenetic conservation values of ORF2p variants used for sensitivity analysis.

The sequences shown in Table 3.9 were aligned using the Geneious multiple protein sequence alignment tool and identity calculation tools. Column one indicates the ORF2p trialanine variant. Columns two and three show the conservation results for mammals only and then for mammals and other vertebrates. Since each variant spans three amino acids, the value shown is that of the residue with the highest identity score. The fourth column shows the average retrotransposition of the given variant. The last column indicates the domain of ORF2p in which the variant lies. For all columns, low is indicated in red and high is indicated in green. The last two columns show the retrotransposition efficiency (green : >80% of WT, gray : 25-80% of WT, and red : <25% of WT) and domain within which each variant is located.

Mutant ID	Conservation (residue with highest identity value as representative for 3xala mutant)		Avg retroT	Domain
	Mammals	Mammals + Verts		
pEA1362_2_3_TG	14	4	25	EN
pEA920_4_6_STS	20	6	72	EN
pEA921_7_9_HIT	62	29	18	EN
pEA922_10_12_ILT	56	52	4	EN
pEA923_13_15_LNI	100	100	6	EN
pEA924_16_18_NGL	100	96	5	EN
pEA925_19_21_NSA	100	67	9	EN
pEA926_22_24_IKR	100	96	11	EN
pEA927_25_27_HRL	86	27	36	EN
pEA928_28_30_ASW	100	14	18	EN
pEA929_31_33_IKS	73	42	10	EN
pEA930_34_36_QDP	100	20	43	EN
pEA931_37_39_SVC	100	66	0	EN
pEA932_40_42_CIQ	100	93	2	EN
pEA933_43_45_ETH	100	100	2	EN
pEA934_46_48_LTC	62	45	6	EN
pEA935_49_51_RDT	100	31	81	EN
pEA936_52_54_HRL	86	68	4	EN
pEA937_55_57_KIK	73	27	15	EN
pEA938_58_60_GWR	100	24	4	EN
pEA939_61_63_KIY	62	37	0	EN
pEA940_64_66_QAN	86	46	84	EN
pEA941_67_69_GKQ	41	13	52	EN
pEA942_70_72_KKA	86	50	7	EN
pEA943_73_75_GVA	100	100	4	EN
pEA944_76_78_ILV	86	60	2	EN
pEA945_79_81_SDK	100	28	2	EN
pEA946_82_84_TDF	86	51	6	EN
pEA947_85_87_KPT	42	14	18	EN
pEA948_88_90_KIK	39	18	10	EN
pEA949_91_93_RDK	100	100	10	EN
pEA950_94_96_EGH	100	100	3	EN
pEA951_97_99_YIM	74	36	2	EN
pEA952_100_102_VKG	100	42	3	EN
pEA953_103_105_SIQ	52	37	28	EN
pEA954_106_108_QEE	60	16	33	EN
pEA955_109_111_LTI	74	38	5	EN
pEA956_112_114_LNI	100	70	2	EN
pEA957_115_117_YAP	100	100	2	EN
pEA958_118_120_NTG	100	70	24	EN
pEA959_121_123_APR	100	15	65	EN
pEA960_124_126_FIK	51	37	5	EN
pEA961_127_129_QVL	100	26	2	EN
pEA962_130_132_SDL	51	38	40	EN
pEA963_133_135_QRD	62	12	23	EN
pEA964_136_138_LDS	56	9	7	EN
pEA965_139_141_HTL	86	64	3	EN
pEA966_142_144_IMG	100	96	2	EN
pEA967_145_147_DFN	100	100	4	EN
pEA968_148_150_TPL	100	39	6	EN
pEA969_151_153_STL	56	24	64	EN
pEA970_154_156_DRS	100	93	2	EN
pEA971_157_159_TRQ	56	17	52	EN
pEA972_160_162_KVN	60	11	48	EN
pEA973_163_165_KDT	62	22	97	EN
pEA974_166_168_QEL	86	31	19	EN
pEA975_169_171_NSA	41	13	79	EN
pEA976_172_174_LHQ	52	17	25	EN
pEA977_175_177_ADL	86	54	8	EN
pEA978_178_180_IDI	100	96	3	EN

Mutant ID	Conservation (residue with highest identity value as representative for 3xala mutant)		Avg retroT	Domain
	Mammals	Mammals + Verts		
pEA979_181_183_YRT	100	86	3	EN
pEA980_184_186_LHP	86	73	5	EN
pEA981_187_189_KST	25	33	88	EN
pEA982_190_192_EYT	86	54	4	EN
pEA983_193_195_FFS	100	93	2	EN
pEA984_196_198_APH	100	76	10	EN
pEA985_199_201_HTY	86	33	6	EN
pEA986_202_204_SKI	100	96	7	EN
pEA987_205_207_DHI	100	100	2	EN
pEA988_208_210_VGS	100	36	7	EN
pEA989_211_213_KAL	73	23	6	EN
pEA990_214_216_LSK	73	21	61	EN
pEA991_217_219_CKR	28	19	92	EN
pEA992_220_222_TEI	86	42	46	EN
pEA993_223_225_ITN	52	31	33	EN
pEA994_226_228_YLS	100	86	17	EN
pEA995_229_231_DHS	100	100	3	EN
pEA996_232_234_AIK	74	43	14	EN
pEA997_235_237_LEL	73	36	7	EN
pEA998_238_240_RIK	51	17	91	EN (1-239)
pEA999_241_243_NLT	15	8	104	DESERT 1 (240-379)
pEA1000_244_246_QSR	21	11	60	DESERT 1
pEA1001_247_249_STT	32	11	88	DESERT 1
pEA1002_250_252_WKL	100	96	5	DESERT 1
pEA1003_253_255_NNL	74	49	42	DESERT 1
pEA1004_256_258_LLN	100	67	10	DESERT 1
pEA1005_259_261_DYW	26	21	29	DESERT 1
pEA1006_262_264_VHN	64	20	106	DESERT 1
pEA1007_265_267_EMK	86	32	22	DESERT 1
pEA1008_268_270_AEI	74	53	13	DESERT 1
pEA1009_271_273_KMF	62	47	7	DESERT 1
pEA1010_274_276_FET	86	23	16	DESERT 1
pEA1011_277_279_NEN	100	96	12	DESERT 1
pEA1012_280_282_KDT	100	28	28	DESERT 1
pEA1013_283_285_TYQ	73	28	19	DESERT 1
pEA1014_286_288_NLW	100	100	3	DESERT 1
pEA1015_289_291_DAF	100	51	3	DESERT 1
pEA1016_292_294_KAV	100	100	3	DESERT 1
pEA1017_295_297_CRG	100	100	5	DESERT 1
pEA1018_298_300_KFI	100	73	3	DESERT 1
pEA1019_301_303_ALN	56	29	103	DESERT 1
pEA1020_304_306_AYK	86	29	73	DESERT 1
pEA1021_307_309_RKQ	86	58	62	DESERT 1
pEA1022_310_312_ERS	73	14	64	DESERT 1
pEA1023_313_315_KID	42	18	62	DESERT 1
pEA1024_316_318_TLT	100	57	59	DESERT 1
pEA1025_319_321_SQL	74	36	73	DESERT 1
pEA1026_322_324_KEL	73	38	69	DESERT 1
pEA1027_325_327_EKQ	100	52	87	DESERT 1
pEA1028_328_330_EQT	46	17	77	DESERT 1
pEA1029_331_333_HSK	86	56	83	DESERT 1
pEA1030_334_336_ASR	74	15	94	DESERT 1
pEA1031_337_339_RQE	86	32	20	DESERT 1
pEA1032_340_342_ITK	100	31	31	DESERT 1
pEA1033_343_345_IRA	74	24	30	DESERT 1
pEA1034_346_348_ELK	100	31	14	DESERT 1
pEA1035_349_351_EIE	100	23	59	DESERT 1
pEA1036_352_354_TQK	39	17	54	DESERT 1
pEA1037_355_357_TLQ	74	25	31	DESERT 1
pEA1038_358_360_KIN	100	28	57	DESERT 1

Mutant ID	Conservation (residue with highest identity value as representative for 3xala mutant)		Avg retroT	Domain
	Mammals	Mammals + Verts		
pEA1039_361_363_ESR	73	40	31	DESERT 1
pEA1040_364_366_SWF	100	25	12	DESERT 1
pEA1041_367_369_FER	100	56	24	DESERT 1
pEA1042_370_372_INK	100	86	17	DESERT 1
pEA1043_373_375_IDR	100	48	61	DESERT 1
pEA1044_376_378_PLA	100	90	42	DESERT 1
pEA1045_379_381_RLI	100	27	65	DESERT 1 (240-379)
pEA1046_382_384_KKK	73	46	57	Z (380-480)
pEA1047_385_387_REK	86	22	59	Z
pEA1048_388_390_NQI	100	80	60	Z
pEA1049_391_393_DTI	100	74	60	Z
pEA1050_394_396_KND	62	22	77	Z
pEA1051_397_399_KGD	86	56	55	Z
pEA1052_400_402_ITT	100	26	22	Z
pEA1053_403_405_DPT	51	27	45	Z
pEA1054_406_408_EIQ	100	96	14	Z
pEA1055_409_411_TTI	86	50	28	Z
pEA1056_412_414_REY	74	50	28	Z
pEA1057_415_417_YKH	74	86	33	Z
pEA1058_418_420_LYA	100	100	31	Z
pEA1059_421_423_NKL	100	21	66	Z
pEA1060_424_426_ENL	100	25	50	Z
pEA1061_427_429_EEM	100	24	38	Z
pEA1062_430_432_DTF	86	51	20	Z
pEA1063_433_435_LDT	86	55	6	Z
pEA1064_436_438_YTL	51	47	18	Z
pEA1065_439_441_PRL	100	54	23	Z
pEA1066_442_444_NQE	60	26	73	Z
pEA1067_445_447_EVE	62	23	43	Z
pEA1068_448_450_SLN	100	80	30	Z
pEA1069_451_453_RPI	100	74	31	Z
pEA1070_454_456_TGS	53	43	88	Z
pEA1071_457_459_EIV	100	96	3	Z
pEA1072_460_462_AII	100	76	16	Z
pEA1073_463_465_NSL	100	53	13	Z
pEA1074_466_468_PTK	100	28	40	Z
pEA1075_469_471_KSP	100	100	62	Z
pEA1076_472_474_GPD	100	100	4	Z
pEA1077_475_477_GFT	100	100	7	Z
pEA1078_478_480_AEF	100	54	18	Z (380-480)
pEA1079_481_483_YQR	100	90	14	
pEA1080_484_486_YKE	86	44	8	
pEA1081_487_489_ELV	74	56	10	
pEA1082_490_492_PFL	100	83	8	
pEA1083_493_495_LKL	86	26	6	
pEA1084_496_498_FQS	100	44	28	RT (498-773)
pEA1085_499_501_IEK	86	24	79	RT
pEA1086_502_504_EGI	100	35	38	RT
pEA1087_505_507_LPN	100	83	8	RT
pEA1088_508_510_SFY	100	41	9	RT
pEA1089_511_513_EAS	86	74	41	RT
pEA1090_514_516_IIL	100	96	6	RT
pEA1091_517_519_IPK	100	100	8	RT
pEA1092_520_522_PGR	100	54	10	RT
pEA1093_523_525_DTT	86	76	80	RT
pEA1094_526_528_KKE	100	27	83	RT
pEA1095_529_531_NFR	100	100	6	RT
pEA1096_532_534_PIS	100	100	6	RT
pEA1097_535_537_LMN	100	89	10	RT
pEA1098_538_540_IDA	100	96	8	RT

Mutant ID	Conservation (residue with highest identity value as representative for 3xala mutant)		Avg retroT	Domain
	Mammals	Mammals + Verts		
pEA1099_541_543_KIL	100	96	4	RT
pEA1100_544_546_NKI	100	67	58	RT
pEA1101_547_549_LAN	100	89	53	RT
pEA1102_550_552_RIQ	100	100	8	RT
pEA1103_553_555_QHI	86	38	89	RT
pEA1104_556_558_KKL	74	34	43	RT
pEA1105_559_561_IHH	100	67	16	RT
pEA1106_562_564_DQV	100	100	6	RT
pEA1107_565_567_GFI	100	100	20	RT
pEA1108_568_570_PGM	100	49	53	RT
pEA1109_571_573_QGW	100	22	16	RT
pEA1110_574_576_FNI	100	90	36	RT
pEA1111_577_579_RKS	100	93	5	RT
pEA1112_580_582_INV	100	30	30	RT
pEA1113_583_585_IQH	100	42	61	RT
pEA1114_586_588_INR	100	14	57	RT
pEA1115_589_591_AKD	60	17	79	RT
pEA1116_592_594_KNH	100	16	54	RT
pEA1117_595_597_MII	100	32	6	RT
pEA1118_598_600_SID	100	100	5	RT
pEA1119_601_603_AEK	100	100	62	RT
pEA1120_604_606_AFD	100	100	4	RT
pEA1121_607_609_KIQ	100	40	39	RT
pEA1122_610_612_QPF	100	53	5	RT
pEA1123_613_615_MLK	86	46	6	RT
pEA1124_616_618_TLN	100	56	35	RT
pEA1125_619_621_KLG	100	46	13	RT
pEA1126_622_624_IDG	100	38	8	RT
pEA1127_625_627_TYF	86	62	19	RT
pEA1128_628_630_KII	73	46	4	RT
pEA1129_631_633_RAI	100	46	25	RT
pEA1130_634_636_YDK	100	96	14	RT
pEA1131_637_639_PTA	100	96	63	RT
pEA1132_640_642_NII	100	48	4	RT
pEA1133_643_645_LNG	100	86	13	RT
pEA1134_646_648_QKL	100	37	45	RT
pEA1135_649_651_EAF	64	55	6	RT
pEA1136_652_654_PLK	86	44	35	RT
pEA1137_655_657_TGT	100	96	11	RT
pEA1138_658_660_RQG	100	100	9	RT
pEA1139_661_663_CPL	100	100	6	RT
pEA1140_664_666_SPL	100	100	5	RT
pEA1141_667_669_LFN	100	93	12	RT
pEA1142_670_672_IVL	100	39	14	RT
pEA1143_673_675_EVL	100	100	5	RT
pEA1144_676_678_ARA	100	53	123	RT
pEA1145_679_681_IRQ	100	77	6	RT
pEA1146_682_684_EKE	60	16	29	RT
pEA1147_685_687_IKG	100	93	4	RT
pEA1148_688_690_IQL	100	39	9	RT
pEA1149_691_693_GKE	100	25	52	RT
pEA1150_694_696_EVK	100	74	3	RT
pEA1151_697_699_LSL	100	58	37	RT
pEA1152_700_702_FAD	100	100	9	RT
pEA1153_703_705_DMI	100	100	4	RT
pEA1154_706_708_VYL	100	39	7	RT
pEA1155_709_711_ENP	100	80	16	RT
pEA1156_712_714_IVS	100	46	34	RT
pEA1157_715_717_AQN	60	25	88	RT
pEA1158_718_720_LLK	100	34	7	RT

Mutant ID	Conservation (residue with highest identity value as representative for 3xala mutant)		Avg retroT	Domain
	Mammals	Mammals + Verts		
pEA1159_721_723_LIS	100	51	6	RT
pEA1160_724_726_NFS	74	56	5	RT
pEA1161_727_729_KVS	100	57	20	RT
pEA1162_730_732_GYK	100	93	6	RT
pEA1163_733_735_INV	100	96	11	RT
pEA1164_736_738_QKS	100	100	5	RT
pEA1165_739_741_QAF	100	33	8	RT
pEA1166_742_744_LYT	100	33	6	RT
pEA1167_745_747_NNR	73	12	60	RT
pEA1168_748_750_QTE	62	12	92	RT
pEA1169_751_753_SQI	62	21	120	RT
pEA1170_754_756_MGE	31	9	116	RT
pEA1171_757_759_LPF	100	34	22	RT
pEA1172_760_762_TIA	86	23	92	RT
pEA1173_763_765_SKR	41	22	66	RT
pEA1174_766_768_IKY	100	100	10	RT
pEA1175_769_771_LGI	100	100	4	RT
pEA1176_772_774_QLT	100	36	20	RT (498-773)
pEA1177_775_777_RDV	64	16	78	DESERT 2 (774-1275)
pEA1178_778_780_KDL	100	52	44	DESERT 2
pEA1179_781_783_FKE	86	19	83	DESERT 2
pEA1180_784_786_NYK	100	100	14	DESERT 2
pEA1181_787_789_PLL	86	30	42	DESERT 2
pEA1182_790_792_KEI	73	37	54	DESERT 2
pEA1183_793_795_KEE	100	35	75	DESERT 2
pEA1184_796_798_TNK	40	32	104	DESERT 2
pEA1185_799_801_WKN	100	100	62	DESERT 2
pEA1186_802_804_IPC	100	47	80	DESERT 2
pEA1187_805_807_SWV	100	65	25	DESERT 2
pEA1188_808_810_GRI	100	77	34	DESERT 2
pEA1189_811_813_NIV	100	35	41	DESERT 2
pEA1190_814_816_KMA	100	100	35	DESERT 2
pEA1191_817_819_ILP	100	96	11	DESERT 2
pEA1192_820_822_KVI	100	64	19	DESERT 2
pEA1193_823_825_YRF	100	80	100	DESERT 2
pEA1194_826_828_NAI	100	35	56	DESERT 2
pEA1195_829_831_PIK	100	96	5	DESERT 2
pEA1196_832_834_LPM	86	42	66	DESERT 2
pEA1197_835_837_TFF	100	73	35	DESERT 2
pEA1198_838_840_TEL	53	29	67	DESERT 2
pEA1199_841_843_EKT	86	23	82	DESERT 2
pEA1200_844_846_TLK	52	31	98	DESERT 2
pEA1201_847_849_FIW	100	89	8	DESERT 2
pEA1202_850_852_NQK	86	70	83	DESERT 2
pEA1203_853_855_RAR	100	86	19	DESERT 2
pEA1204_856_858_IAK	100	60	52	DESERT 2
pEA1205_859_861_SIL	86	68	56	DESERT 2
pEA1206_862_864_SQK	86	19	73	DESERT 2
pEA1207_865_867_NKA	42	24	122	DESERT 2
pEA1208_868_870_GGI	100	100	60	DESERT 2
pEA1209_871_873_TLP	100	100	69	DESERT 2
pEA1210_874_876_DFK	86	38	7	DESERT 2
pEA1211_877_879_LYY	100	100	6	DESERT 2
pEA1212_880_882_KAT	100	83	60	DESERT 2
pEA1213_883_885_VTK	100	34	62	DESERT 2
pEA1214_886_888_TAW	100	20	36	DESERT 2
pEA1215_889_891_YWY	100	70	4	DESERT 2
pEA1216_892_894_QNR	100	13	38	DESERT 2
pEA1217_895_897_DID	100	12	55	DESERT 2
pEA1218_898_900_QWN	100	49	60	DESERT 2

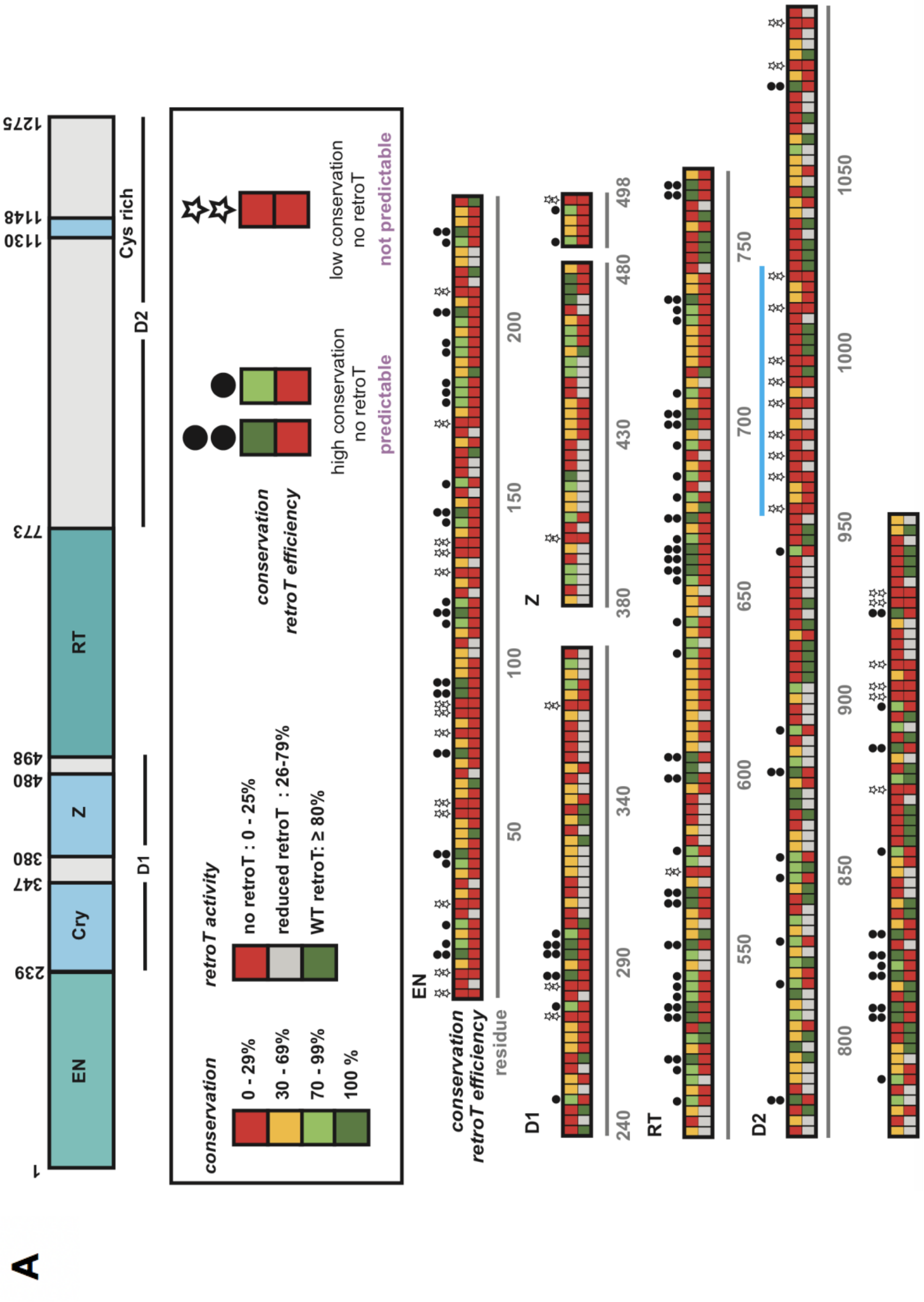
Mutant ID	Conservation (residue with highest identity value as representative for 3xala mutant)		Avg retroT	Domain
	Mammals	Mammals + Verts		
pEA1219_901_903_RTE	86	83	69	DESERT 2
pEA1220_904_906_PSE	64	13	120	DESERT 2
pEA1221_907_909_IMP	86	19	112	DESERT 2
pEA1222_910_912_HIY	73	19	88	DESERT 2
pEA1223_913_915_NYL	74	26	105	DESERT 2
pEA1224_916_918_IFD	100	34	7	DESERT 2
pEA1225_919_921_KPE	86	19	95	DESERT 2
pEA1226_922_924_KNK	60	20	79	DESERT 2
pEA1227_925_927_QWG	86	9	112	DESERT 2
pEA1228_928_930_KDS	86	18	73	DESERT 2
pEA1229_931_933_LFN	100	22	49	DESERT 2
pEA1230_934_936_KWC	100	27	75	DESERT 2
pEA1231_937_939_WEN	86	19	64	DESERT 2
pEA1232_940_942_WLA	100	79	19	DESERT 2
pEA1233_943_945_ICR	86	16	96	DESERT 2
pEA1234_946_948_KLK	51	21	97	DESERT 2
pEA1235_949_951_LDP	86	14	71	DESERT 2
pEA1236_952_954_FLT	86	23	16	DESERT 2
pEA1237_955_957_PYT	100	40	4	DESERT 2
pEA1238_958_960_KIN	100	46	8	DESERT 2
pEA1239_961_963_SRW	100	30	10	DESERT 2
pEA1240_964_966_IKD	86	22	42	DESERT 2
pEA1241_967_969_LNV	100	16	8	DESERT 2
pEA1242_970_972_KPK	47	13	64	DESERT 2
pEA1243_973_975_TIK	74	16	14	DESERT 2
pEA1244_976_978_TLE	60	47	28	DESERT 2
pEA1245_979_981_ENL	62	22	63	DESERT 2
pEA1246_982_984_GIT	86	22	25	DESERT 2
pEA1247_985_987_IQD	51	34	51	DESERT 2
pEA1248_988_990_IGV	35	29	10	DESERT 2
pEA1249_991_993_GKD	34	14	97	DESERT 2
pEA1250_994_996_FMS	53	30	8	DESERT 2
pEA1251_997_999_KTP	31	27	93	DESERT 2
pEA1252_1000_1002_KAM	40	21	103	DESERT 2
pEA1253_1003_1005_ATK	28	15	109	DESERT 2
pEA1254_1006_1008_DKI	53	20	77	DESERT 2
pEA1255_1009_1011_DKW	86	23	11	DESERT 2
pEA1256_1012_1014_DLI	74	45	11	DESERT 2
pEA1257_1015_1017_KLK	53	48	11	DESERT 2
pEA1258_1018_1020_SFC	86	21	6	DESERT 2
pEA1259_1021_1023_TAK	86	17	93	DESERT 2
pEA1260_1024_1026_ETT	42	9	91	DESERT 2
pEA1261_1027_1029_IRV	42	8	110	DESERT 2
pEA1262_1030_1032_NRQ	86	14	72	DESERT 2
pEA1263_1033_1035_PTT	86	24	126	DESERT 2
pEA1264_1036_1038_WEK	100	42	29	DESERT 2
pEA1265_1039_1041_IFA	100	34	8	DESERT 2
pEA1266_1042_1044_TYS	26	10	106	DESERT 2
pEA1267_1045_1047_SDK	100	42	41	DESERT 2
pEA1268_1048_1050_GLI	100	44	15	DESERT 2
pEA1269_1051_1053_SRI	100	60	55	DESERT 2
pEA1270_1054_1056_YNE	100	96	66	DESERT 2
pEA1271_1057_1059_LKQ	86	44	82	DESERT 2
pEA1272_1060_1062_IYK	53	12	67	DESERT 2
pEA1273_1063_1065_KKT	53	12	108	DESERT 2
pEA1274_1066_1068_NNP	86	14	57	DESERT 2
pEA1275_1069_1071_IKK	74	29	43	DESERT 2
pEA1276_1072_1074_WAK	100	100	18	DESERT 2
pEA1277_1075_1077_DMN	74	35	5	DESERT 2
pEA1278_1078_1080_RHF	86	27	14	DESERT 2
pEA1279_1081_1083_SKE	74	35	94	DESERT 2
pEA1280_1084_1086_DIY	50	51	65	DESERT 2

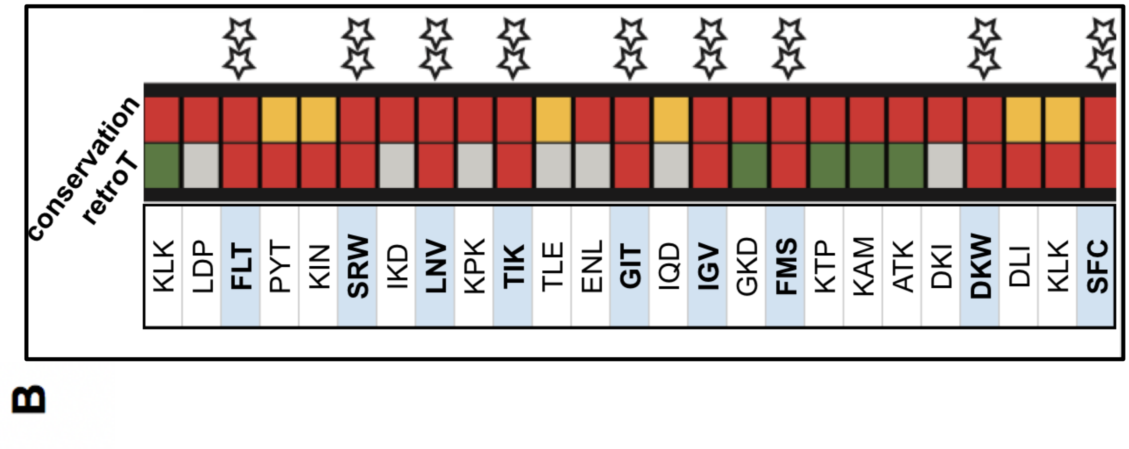
Mutant ID	Conservation (residue with highest identity value as representative for 3xala mutant)		Avg retroT	Domain
	Mammals	Mammals + Verts		
pEA1281_1087_1089_AAK	86	21	64	DESERT 2
pEA1282_1090_1092_KHM	86	14	15	DESERT 2
pEA1283_1093_1095_KKC	100	24	65	DESERT 2
pEA1284_1096_1098_SSS	100	33	44	DESERT 2
pEA1285_1099_1101_LAI	100	20	50	DESERT 2
pEA1286_1102_1104_REM	100	41	24	DESERT 2
pEA1287_1105_1107_QIK	100	89	28	DESERT 2
pEA1288_1108_1110_TTM	100	24	44	DESERT 2
pEA1289_1111_1113_RYH	100	70	5	DESERT 2
pEA1290_1114_1116_LTP	100	69	21	DESERT 2
pEA1291_1117_1119_VRM	100	55	70	DESERT 2
pEA1292_1120_1122_AII	86	33	117	DESERT 2
pEA1293_1123_1125_KKS	62	27	93	DESERT 2
pEA1294_1126_1128_GNN	51	17	97	DESERT 2
pEA1295_1129_1131_RCW	100	100	8	DESERT 2
pEA1296_1132_1134_RGC	100	100	6	DESERT 2
pEA1297_1135_1137_GEI	86	19	127	DESERT 2
pEA1298_1138_1140_GTL	100	49	82	DESERT 2
pEA1299_1141_1143_LHC	100	100	5	DESERT 2
pEA1300_1144_1146_WWD	100	89	19	DESERT 2
pEA1301_1147_1149_CKL	100	100	11	DESERT 2
pEA1302_1150_1152_VQP	100	26	46	DESERT 2
pEA1303_1153_1155_LWK	100	100	11	DESERT 2
pEA1304_1156_1158_SVW	100	44	24	DESERT 2
pEA1305_1159_1161_RFL	86	29	87	DESERT 2
pEA1306_1162_1164_RDL	62	45	100	DESERT 2
pEA1307_1165_1167_ELE	60	16	79	DESERT 2
pEA1308_1168_1170_IPF	100	27	50	DESERT 2
pEA1309_1171_1173_DPA	100	67	70	DESERT 2
pEA1310_1174_1176_IPL	100	61	21	DESERT 2
pEA1311_1177_1179_LGI	100	71	4	DESERT 2
pEA1312_1180_1182_YPN	86	17	87	DESERT 2
pEA1313_1183_1185_EYK	34	10	109	DESERT 2
pEA1314_1186_1188_SCC	15	11	97	DESERT 2
pEA1315_1189_1191_YKD	51	13	94	DESERT 2
pEA1316_1192_1194_TCT	100	18	64	DESERT 2
pEA1317_1195_1197_RMF	100	18	19	DESERT 2
pEA1318_1198_1200_IAA	100	86	102	DESERT 2
pEA1319_1201_1203_LFT	62	26	82	DESERT 2
pEA1320_1204_1206_IAK	100	52	68	DESERT 2
pEA1321_1207_1209_TWN	100	100	13	DESERT 2
pEA1322_1210_1212_QPK	86	20	86	DESERT 2
pEA1323_1213_1215_CPT	100	73	63	DESERT 2
pEA1324_1216_1218_MID	51	16	91	DESERT 2
pEA1325_1219_1221_WIK	100	89	25	DESERT 2
pEA1326_1222_1224_KMW	100	21	8	DESERT 2
pEA1327_1225_1227_HIY	62	22	24	DESERT 2
pEA1328_1228_1230_TME	100	51	5	DESERT 2
pEA1329_1231_1233_YYA	100	23	23	DESERT 2
pEA1330_1234_1236_AIK	73	22	51	DESERT 2
pEA1331_1237_1239_NDE	46	17	61	DESERT 2
pEA1332_1240_1242_FIS	19	15	60	DESERT 2
pEA1333_1243_1245_FVG	43	37	28	DESERT 2
pEA1334_1246_1248_TWM	100	100	16	DESERT 2
pEA1335_1249_1251_KLE	100	22	21	DESERT 2
pEA1336_1252_1254_TII	62	19	9	DESERT 2
pEA1337_1255_1257_LSK	100	14	56	DESERT 2
pEA1338_1258_1260_LSQ	73	12	82	DESERT 2
pEA1339_1261_1263_EQK	64	15	104	DESERT 2
pEA1340_1264_1266_TKH	42	20	102	DESERT 2
pEA1341_1267_1269_RIF	46	20	49	DESERT 2
pEA1342_1270_1272_SLI	51	35	81	DESERT 2
pEA1343_1273_1275_GGN	51	40	58	DESERT 2

Figure 3.19 : Trends in amino acid conservation and sensitivity to mutation across ORF2p.

(A)The schematic for the ORF2p domains is along the top. This is a graphical representation and interpretation of the data displayed in Table 3.10 (mammals and all vertebrates, so the mammals alone column is excluded). For each trialanine variant, we took the value that corresponded to the residue that had the highest conservation to represent the mutant. As shown in the box (low being red and high being green), the conservation and retrotransposition are color-coded. They are then stacked to show how conservation and activity compare. One dot and two black dots above a mutant mean that, as expected, there was strong conservation and no L1 jumping. Stars below the mutant mean that there is low conservation and no retrotransposition, highlighting areas that may be important in ORF2p that were not predicted by conservation alone. The *Star Cluster* region is indicated with a light blue bar, which is shown (B) zoomed in and in detail with the three WT amino acids (in single letter format) corresponding to each variant.

Figure 3.19 spans the next 2 pages.


A
1
239
347
380
480
498
773
1130
1148
1275
EN
Cry
Z
RT
D1
D2
Cys rich
conservation
0 - 29%
30 - 69%
70 - 99%
100 %
retroT activity
no retroT : 0 - 25%
reduced retroT : 26-79%
WT retroT: ≥ 80%
conservation
retroT efficiency
high conservation
no retroT
predictable
low conservation
no retroT
not predictable
EN
conservation
retroT efficiency
residue
50
100
150
200
D1
240
290
340
Z
380
430
480
498
RT
550
600
650
700
750
D2
800
850
900
950
1000
1050


B
conservation
retroT
KLK
LDP
FLT
PYT
KIN
SRW
IKD
LNV
KPK
TIK
TLE
ENL
GIT
IQD
IGV
GKD
FMS
KTP
KAM
ATK
DKI
DKW
DLI
KLK
SFC

ORF2p has large segments that have yet to be well-characterized. What are the most important motifs in ORF2p that should be investigated next? This work helps us reach beyond the most studied regions of ORF2 and creates a framework for prioritizing functional regions for further study. As shown, we have found regions of ORF2p that have not experienced purifying selection and are not well conserved, but are extremely sensitive to mutation, which could not have been predicted. The motifs that are also essential interest are those that coincide with the 12% of ORF2p variants that show activity reduced to <5% that of WT (and subsequently the additional ~43% of ORF2p variants with activity reduced to <25% that of WT). Regions we should study also include motifs that are conserved and have impaired retrotransposition. This helps orient us to pointedly ask what motifs in the ORF2 protein are most important in host-parasite cell biology and how they are involved in supporting L1 retrotransposition.

Methods

Design of the trialanine scanning mutagenic library

A major goal was to create a pipeline in which an ordered (as opposed to pooled) library could be efficiently assembled. We worked from well-established constructs in the lab that we knew transfected well and supported high retrotransposition frequency. The original vector backbone that has been continuously re-engineered in our lab is the pCEP4 vector (ThermoFisher pCEP4 Catalog no. V044-50), which we refer to as pCEP-puro (the original HygroR cassette was replaced with a PuroR cassette). This is the basic backbone of the parental plasmid, into which each trialanine variant was cloned into a WT version of human L1, pEA0264, serving as the "Destination vector" shown in Figure 3.3. We added a KanR cassette (knowing that all of the mutant DNA would arrive in plasmids with CarbR cassettes because we needed to be able to select for the destination vector). pEA0264 contains a human L1-rp cassette, under the TET (inducible, minimal-CMV) promoter. There is no native L1 5'UTR sequence. The full native 3'UTR sequence is present, but interrupted by the GFP-AI fluorescent retrotransposition reporter construct [78]. Although the L1 native 3'UTR has a weak polyA addition signal, there is also the SV40 polyA addition signal from pCEP4. For a measure of transfection efficiency, we also introduced a constitutively expressed fluorogen-activating protein [79,80], FAP-DL5 (this is not indicated in the figure, nor was it used in this study).

As described in the text, unique restriction sites were designed such that they fell only within L1 and not in the vector backbone, and spaced roughly equally about every 600bp. This entailed both removing and adding (silently, if in a coding region) restriction enzyme cut sites from throughout the plasmid backbone and the L1-rp cassette (using the

GeneDesign online tool [81]). Plasmids were tested at every step in cloning sites in and out of the plasmid to be sure that there was no impact on retrotransposition efficiency relative to WT. Comparing the primary sequence of L1-rp (accession number AF148856) to the WT-L1 cassette of pEA0264 there are the following changes : 14bp in the coding region, 2bp in the ORF2 stop codon, 3bp immediately after the ORF2 stop codon, and 7bp in the inter-ORF region (which were changes that had been made in cassettes that pre-date any pEA constructs). For pEA0264, some cut sites remained unique in the L1 cassette for cloning purposes with pEA264 (AflII, AgeI, BamHI, BsiWI, BstBI, ClaI, NotI, SbfI, SphI, VspI) and some remained "forbidden" (BstZ17I), for cloning purposes with synthetic mutant DNA plasmids as well as to be strategic in the design of the subsequent Gibson assembly and transformation steps (explained below). Also, for the synthetic mutant fragment DNAs that were ordered, forbidden sites included BsaI or AarI sites as well the sites were to remain unique in the L1 cassette. This was mainly to facilitate downstream combinatorial cloning or manipulation of the individual mutants much easier down the line. Thus, none of the "unique L1" sites were duplicated in any of the 538 mutant constructs, but, unavoidably, the trialanine variants that overlap with the unique sites must remove them. The trialanine 9bp sequence was also designed to lack CGs (to avoid creating CpG methylation) and to contain a PstI site. PstI cuts WT pEA0264 into 4 pieces, that separate well on an agarose gel; PstI cutting the trialanine DNA sequenced allowed for detecting correct assemblies in a high throughput and cost-effective manner for most variants (Figure 3.5).

Build and validation of the library of 538 trialanine mutants

The 538 trialanine mutants were generated using Gibson assembly [82], as shown in Figure 3.3. Each variant was contained within 1 of 9 chunks, each chunk's worth of final library clones was made as a unit in a 96-well plate format. All enzymes used for the library build are listed here : BstZ17I-HF (20 U/μL, NEB cat # R3594), PstI-HF (20 U/μL, NEB cat # R3140) , Alkaline Phosphatase (CIP) (10 U/μL, NEB cat # M0290), T5 Exonuclease (10 U/μL, NEB cat # M0363), Phusion High-Fidelity DNA Polymerase (2 U/μL, NEB cat # M0530), Taq DNA Ligase (40 U/μL, NEB cat # M0208), NotI-HF (20 U/μL, NEB cat # R3189), SbfI (10 U/μL, NEB cat # R0642), AgeI-HF (20 U/μL, NEB cat # R3552), AseI (10 U/μL, NEB cat # R0526), ClaI (5 U/μL, NEB cat # R0197), AflII (20 U/μL, NEB cat # R0520), BsiWI-HF (20 U/μL, NEB cat # R3553) , BamHI-HF (20 U/μL, NEB cat # R3136), BstBI (20 U/μL, NEB cat # R0519), SphI (10 U/μL, NEB cat # R0182).

Digest and purify the pEA0264 backbone :

All constructs for assembly were digested so that the appropriate backbone and insert were linearized. First, the backbone was prepared for each of the 9 chunks (see Table 3.1) by digesting pEA0264 into nine separate reactions with the appropriate two restrictions enzymes. One digestion reaction contained 20 μg pEA0264, 50 U of calf intestinal phosphatase (CIP), 60 U each restriction enzyme, with the appropriate enzyme buffer and ddH_2O to a final volume of 220 μL and incubated at the appropriate temperature for three hours. Approximately 2μg of the final reaction was ran per lane of a 1% agarose gel and the band containing the digested pEA264 backbone fragment (with the WT chunk DNA excised) DNA was cleaned from the gel using the Zymoclean™ Gel DNA Recovery Kit (Zymogen, cat. # D4002).

Digest the synthetic DNA-containing plasmids :

The DNA for the synthesized mutant fragments (received through special order mainly from Gen9 and some from Qinglan) was organized in nine 96-well plates, corresponding to each chunk. Multichannel pipetting with filter tips was used for all steps throughout the library build. Each mutant-fragment-containing plasmid was designed such that the ~600bp trialanine DNA fragment would be uniquely cut out with BstZ17I, an enzyme that does not cut pEA0264 or any trialanine variant. Thus, these restriction digests did not need to be heat inactivated or cleaned from a gel. Each digest contained 50 ng of plasmid, 1 μL 10x CutSmart Buffer (NEB), 0.5 U BstZ17I-HF, and ddH_2O to a final volume of 10 μL and incubated at 37°C for 1.5 hours.

2.5 μL 2-piece Gibson assembly reactions :

We then used the 96-well approach to performing the Gibson reactions and transformations (as outlined in Figures 3.3 (bottom) and 3.4When prepared fresh concentrated Gibson Master Mix solutions and kept them on ice. These contained all enzymes for the Gibson reaction as well as the appropriate cut pEA0264 product, such that each cut mutant fragment (still in the digestion reaction solution) product could be added to the Master Mix for a final 1x mixture. Each Gibson reaction contained 0.5 μL ISO Buffer [82], 0.001 μL T5 Exo, 0.03 μL Phusion Polymerase (2 U/μL), 0.25 μL Taq DNA Ligase (40 U/μL), 0.3 μL backbone DNA (125ng total), 0.39 μL ddH_2O (for a total of 1.5 μL Master Mix) and 2 μL of the reaction containing the cut DNA fragment, corresponding to 5ng (total) of the ~600bp mutant DNA insert. Reaction were mixed, gently spun down, and incubated for 30 min at 50 °C.

Transformation and plating for Gibson assembly products :

The plate of Gibson reactions was then incubated on a metal block on ice to cool down for 15 min. For the transformation, 25 μL chemically competent DH5α *E. Coli* cells were then added directly to each of the wells containing the 2.5 μL Gibson assembly, incubated on a metal block on ice for 30 minutes, heat-shocked on a metal block at 42°C for 45 seconds. We added 225 μL Luria-Bertani (LB) broth for the out-growth step in which the plate was incubated, non-agitated (covered) at 37°C for 1 hour. The plate was then spun down at 3,000 RPM for 5 minutes to pellet the cells. 170 μL of the supernatant was then removed from each well. All of the transformed cells were in a in a total the ~80 μL LB. The cells in each well were then resuspended (by pipetting up and down 10 times) that remained in the well.

We plated on LB-Kanamycin (Kan) agar plates (100 μg/mL Kan, in petri dishes). (Kan was strategically used for selection so that anything that remained circular in the BstZ17I-digest of the mutant plasmids could not grow; pEA0264 has the KanR cassette). We plated 20% of the cells in 20 μL using what we term the "drop method". 8 transformation mixtures can be plated per plate. Transformations are multichannel pipetted slowly onto a plate to create 8 big droplets, then the plate was slowly moved to a 45° angle to the bench to let the drops flow in parallel paths down the plate (careful not to touch the edges, nor one another). The plates were allowed to dry at room temperature for about 10 minutes and then incubated upside-down at 30°C for 16-24 hours. The lanes of transformant colonies formed using this method are shown in the insert of Figure 3.4.

Colony picking, quality control, and transfection-quality DNA preparation :

After familiarizing ourselves with the efficiency of the assemblies (data not shown), we knew that at least 95% of the colonies were correct for each chunk, and we thus picked only one colony per Gibson assembly plated transformation product (Figure 3.4) A single colony was put into two samples of growth media: (i) to grow 96-well cultures, for low-quality DNA preparation, suitable for isolating plasmid and checking the integrity and correctness of the assembly by PstI digest and (ii) to grow a 5 mL culture in order to prepare a large amount of high-quality (endotoxin-free) DNA, suitable for 96-well transfection into HeLa cells. Thus, one colony was dipped into media (LB- 100 μg/mL Kan) twice: \ (i) into in a well of a 96-

well deep-well containing a glass bead and 1110 μl of media and for \ (ii) into 5mL media in a glass tube. All were grown for 16 hours at 30°C, (i) in an incubator with a shaker or (ii) in a roller drum.

Customized 96-well miniprep and clone check by PstI digest :

We performed a 96-well miniprep using buffers from Qiagen. We harvested the cells grown overnight by centrifuging cells at 3,500 RPM for 10 min, discarded the supernatant, resuspend cells in 250 μl of P1 buffer (Qiagen) by pipetting up and down 10 times, added 250 μl of P2 buffer (Lysis buffer; Qiagen), sealed the plate and mixed by inverting the plate 8 times, unsealed the plate, added 350 μl N3 buffer (neutralization buffer; Qiagen), sealed the plate and mixed by inverting the plate 8 times, transferred the cell lysate (800 μl) to clearing stack (Thermo #278011) on top of binding plate (Promega #A2271) on top of collection plate, and centrifuged for 5 min at 3,000 RPM. We discarded flow through in the collection plate and remove filtrate plate and added 200 μl of PB buffer (Qiagen) to the binding plate, centrifuged for 5 min at 3,000 RPM and discarded flow through. We then added 400 μl of PE buffer (Qiagen) to the binding plate and centrifuge for 5 min at 3000 RPM and discarded flow through. We centrifuged again for 30 min at 3,000 RPM to remove residual ethanol on binding plate. We then transferred the binding plate and placed it on top of a clear bottom plate for elution. We added 100 μl of H_2O and to the binding plate and centrifuged for 5 min at 3,000 RPM to elute the DNA. The overall concentration of the DNAs was estimated by measuring six random samples throughout the plate by measurements of concentration on a Nanodrop UV spectrophotometer.

Roughly 600 ng of the purified plasmid DNA (and the pEA0264 WT reference) was digested in a separate 96-well plate with 4 units of PstI in 10 μL at 37°C for 60 min, run on a

1 % agarose gel and checked for backbone integrity and the correct mutant banding pattern. An example of this quality control assay in shown in Figure 3.5.

Preparation of high-quality DNA and Sanger sequencing :

Each 5 mL saturated culture for each candidate clone (assuming it looked correct by PstI digest) was prepped for high-quality DNA for transfection into Hela cells. We used the ZymoPURE™ Plasmid Miniprep Kit (Zymogen Research, cat. #D4212) and followed the manual's instructions and eluted in 30 µL. The DNA concentration of each construct was measured by Nanodrop and normalized to 100 ng/µL and was stored in a 1.5mL microtube at –20°C. The majority of the volume was then moved to 96-well plates, one for each chunk (to make transfections easier down the line). We Sanger sequenced each clone in from these 96-well plates (to avoid accidental sample mix-up, we knew that what we sequenced were the plasmids in the well they would be transfected from, unless incorrect). We sequenced the entire span of the synthesized fragment that was inserted into pEA0264 as well as through (and past) the Gibson homology arms that were used. Two primers were used to validate each clone.

If a clone was incorrect by either quality control step, we picked a new clone. If that was incorrect again, we redid the Gibson assembly (this was very rare). Once we knew that each clone was correct by digest and Sanger sequencing and had high-quality DNA in a 96-well plate, that trialanine variant was considered built.

96-well retrotransposition assay

Retrotransposition was measured as outlined in Figure 3.7. HeLa-M2 cells [83] were cultured in DMEM media with 10% FBS (Gemini, prod. number 100–106), Penicillin-Streptomycin (stock contains 10,000 units/mL of penicillin and 10,000 µg/mL of streptomycin and is diluted 1:100 in our media; Thermofisher, cat #15140122), and 1 mM L-glutamine (ThermoFisher/Life Technologies, prod. number 25030–081), which will be referred to as just DMEM in the Methods, never growing more than 90% confluent and passaged routinely until seeded for this assay. All of the following was done using 96-well plates and multichannel pipettes. Three chunks were tested in each experiment, two plates each.

On day one, 25,000 HeLa cells were seeded per well in 50 µL DMEM. About an hour later the DNA was transfected. One well of a transfection got the following mixture : 0.2 µL Fugene-HD (Promega, cat #E3211), 60 ng DNA (0.6µL 100ng/ µL stock), 10 µL Opti-mem total. We made enough mixture for 2.5 transfections (since each transfection mix needed to have 2 separate transfections). This incubated at room temperature for 25 minutes. Each well got 10 µL of the transfection mixture and all was pipetted up and down 6 times. 24 hours later, puromycin (puro) was added to each well in 50 µL (2 µg/mL) giving a final concentration of 1 µg/mL. 24 hours later, the cells were split to a black walled tissue culture plate and doxycycline (dox) was added. This was done 2 plates at a time. The final condition for each well was 20% of the cells in 120 µL total ~DMEM, 1 µg/mL puro.

First, we added 100 µL DMEM (with 1.2x dox and 1.2x puro) to each well on the new plates. Each well had 20% of the cells transferred in 20 µL. One the destination plate for the cells was ready the cells in the culture plates were prepared for the split. We removed media from the wells, washed cells once with 50 µL PBS, aspirated off the PBS, add 50 µL trypsinization solution (TrypLE™ Express Enzyme (1X), Life Technologies cat # 12604039), incubated the plates for 5 minutes in the 37°C incubator, then pipetted up and down 3 times to remove clumps in 40 µL to avoid bubbles. Then we added 50 µL DMEM and pipetted up and down 3 times. Then, 20 µL cells were pipetted to respective wells. It was very important to slowly pipette the cells into the new wells, covering the whole surface evenly without pipetting up and down. The cells needed to grow spread out without clumping. These plates then incubated for 3 days.

After incubation the cells were fixed and stained for analysis. First each well of cells was prefixed in 15μL 1% formalin solution by adding 11% formaldehyde in PBS (3mL 37% formaldehyde in 7mL PBS), then incubated at room temperature for 10 minutes, then media was aspirated off, then the cells were fixed in 100 μL 4% formalin solution by adding 4% formaldehyde in PBS (1mL 37% formaldehyde in 9mL PBS), and then incubated at room temperature for 10 minutes, and finally the solution was aspirated off. Then the cells were washed 1 time with 100 μL 0.1% TritonX-100 in PBS-glycine (0.1% TritonX-100, 1xPBS, 10 mM Glycine) and incubated for 2 minutes. Then 2 more 2-minute washes with 100 μL PBS. Then the cells were stained with DAPI (FluoroPure™ grade DAPI, ThermoFisher cat. #D21490, diluted 15 μL in 12 mL PBS) for 45 minutes at room temperature Then this stain was aspirated off, 150 μL PBS was added to each well, and the plates were sealed and imaged at the NYU High Throughput Biology Core for data analysis, discussed below.

Quantification of retotransposition for all mutants

Dr. David Kahler of the High Throughput Biology Core used Arrayscan VTI software for image acquisition and conducted analysis using the Target Activation Bioapplication (Thermo Scientific Cellomics Scan version 6.6.0 (Build 8153) for a 96 well assay. He used the following paramteres: 5x magnification, 2x2 binning 4 fields per well. DAPI positive nuclei were identified using the dynamic isodata thresholding algorithm after minimal background subtraction. DAPI images were used to identify cell nuclei and to delineate the nuclear edges. A ‘circle’ (x= 2 μm greater than the nuclei border was used to identify cells expressing cytoplasmic GFP). Limits of fluorescence were set so that no cells were considered positive for preparations of cells not containing GFP. The reported parameters are explained as follows : Total = total number of DAPI nuclei counted; GFP+ = above GFP threshold; (GFP+/Total*100)mutant / (GFP+/Total*100)WT = retrotransposition efficiency.

Western blot analysis for ORF1 protein

Each ORF1p variant was tested for protein production. The same HeLa cells cultured in DMEM were used. All the same reagents were used as for the 96 well tissue culture described above. On day one, cells were seeded in a 6-well tissue culture plate (Falcon® 6 Well Clear Flat Bottom TC-Treated Multiwell Cell Culture Plate, cat #353046), 250,000 cells in 2mL per well. 24 hours later, DNA was transfected : 100 μL Opti-mem, 1 μg miniprepped DNA, and 3 μL of Fugene-HD mixed for each construct to be tested and incubated for 25 minutes at room temperature, then added drop wise, evenly around the well. 24 hours later the cells were split such that 20% of the cells were moved to another 6-well plate to prevent over growth. The media was aspirated off, the cells were washed with 2 mL PBS, the PBS was aspirated off and the cells treated with 400 μL TrypLE and incubated for 5 minutes in the 37°C incubator. Then each plate was jarred on the side gently to help completely dislodge the cells from the bottom of the plates and 400 μL DMEM was added to each well. Upon transfer of cells, each well was pipetted up and down to mix and remove clumps. Then 160 μL (20%) was transferred to a new well with 2 mL DMEM, final concentration 1 μg/mL puro. The cells were then incubated for three days and dox was added to a final concentration of 1 μg/mL in 100 μL (final approximate volume 2.3 mL). 24 hours later, the cells were then harvested. We aspirated the media from each well, added 500 μL TrypLE, incubated at 37°C for 5 minutes, jarred on the side to dislodge cells, and 500 μL DMEM was added to each well. Then all cells were transferred to a 1.5 mL microcentrifuge tube and spun at 1,000 RPM for 3 minutes. The supernatant was then aspirated off.

The samples were kept on ice and resuspended in 150 µL of ice-cold Extraction Buffer (500 mM NaCl, 20 mM HEPES, 1% TritonX, and a freshly added protease inhibitor (cOmplete™, EDTA-free Protease Inhibitor Cocktail, Sigma-Aldrich cat #11873580001; one tablet per 50 mL 1x buffer). The samples were then vortexed for 10 seconds and incubated on ice for 10 minutes. We then spun the samples at max speed (14,000 RPM) for 10 minutes at 4°C to pellet insoluble. To prepare these samples for the denaturing gel, 150 µL of the clarified lysate was then moved to a new tube and mixed with 15 µL of 100 mM DTT, 55 µL 4x LDS (NuPAGE® LDS Sample Buffer , Life Technologies, cat. # NP0007), was mixed well, incubated for 10 minutes at 70°C and then either allowed to cool and loaded on a gel or stored at -20°C. 20 µL of each sample was loaded per gel lane (NuPAGE® 4-12% Bis-Tris Midi Protein Gels, Thermofisher cat. #WG1403) and ran in MOPS buffer. Proteins were transferred on Immobilon-FL membrane (Millipore, prod. number IPFL00010), blocked for 1 hour with blocking buffer (LiCOR prod. number 927–40000):TBS buffer (50 mM Tris Base, 154 mM NaCl) 1:1 and then incubated with primary antibodies (α-Tubulin rabbit α/β-Tubulin Antibody from Cell Signaling Technology cat. #2148 diluted 1:1000 and α-ORF1 mouse ORF1 Milliprep, prod. number MABC1152 diluted 1:50,000) and

solubilized in LiCOR blocking buffer:TBS-Tween (0.1% Tween in TBS buffer) 1:1. Secondary donkey anti-goat antibodies conjugated to IRDye680 (anti-rabbit) or IRDye800 (anti-mouse) dyes (LiCOR prod. number 926– 32210 and 926–68071), were used for detection of the specific bands on an Odyssey CLx scanner (LiCOR).

Sequencing : sample preparation and acquiring data

Preparation of lysate from transfects HeLa cells :

Samples for sequencing were always prepared from a 75-90% confluent 10 cm tissue culture plate of HeLa cells that had been transfected (as described above) with the constructs of interest, under puro selection for 5 days and dox induction for 24 hours. We aspirated the media from each well, added 2mL TrypLE, incubated at 37°C for 5 minutes, hit on the side to dislodge cells, and added 2 mL DMEM was added to each plate. Then all cells were transferred to a 15 mL conical tube and spun at 1,000 RPM for 3 min at 4°C. The supernatant was then aspirated off, the cells were resuspended in 1 mL PBS to wash them and move them to a 1.5 mL microcentrifuge tube and spun at 1,000 RPM for 3 min at 4°C. Then the supernatant was aspirated off. Cells were either flash frozen in liquid nitrogen and stored at -80°C or continued into the next steps to prepare lysate. One microcentrifuge tube always corresponded to roughly 1 confluent 10cm plate at this point. Depending on downstream applications, sometimes lysates would be combined to have enough material. The samples were kept on ice and resuspended in 500 μL of ice-cold Extraction Buffer (500 mM NaCl, 20 mM HEPES, 1% TritonX, 40 U/ μL RNAsin (Recombinant RNasin® Ribonuclease Inhibitor, Promega cat. #N2515) and a freshly added protease inhibitor [one Roche Complete tablet per 50 mL). The samples were then vortexed for 10 seconds and incubated on ice for 10 minutes. We then spun the samples at max speed (14,000 RPM) for 10 minutes at 4°C to pellet insoluble and kept the soluble fraction as the clarified lysate for sequencing experiments.

It is important to note that we always checked for expression of ORF1p (and if doing an IP, the efficiency IP of ORF1p) in every experiment by Western analysis before doing any sequencing library preparation or sequencing runs.

Preparation of total plasmid DNA and the sample for DNA sequencing :

To prepare the total plasmid DNA, 500-600 μL of the clarified lysate prepared went through a miniprep treatment (Zyppy™ Plasmid Miniprep Kit, Zymo research, cat. # D4037) in which we followed the manufacturer's protocol and eluted in 30 μL. We the concentrated the DNA (DNA Clean & Concentrator-25, Zymo research, cat. # D4033) in which we followed the manufacturer's protocol and eluted in 6 μL H_2O. We then electroporated 3 μL DNA into 30 μL ElectroMAX™ DH10B™ T1 Phage-Resistant Competent Cells (which had been thawed on ice for ten minutes) in ice cold cuvettes (0.1 cm electroporation cuvettes) using the Gene Pulser®/MicroPulser™. We immediately added 950 μL SOC (that had been incubated at 42°C) and each outgrowth was done in a 1.5 mL microcentrifuge tube at 37°C for one hr. The culture was both plated and put into liquid medium because between 1 and 14 plasmids would be in each pool of plasmid DNA. We plated 5 μL of the transformation (in 100 μL + Carb 75 μg /mL) on LB-Carb plates to ensure the efficiency of each transformation yielded enough colonies to represent the plasmid (sometimes plasmid pool) composition well; this meant, when plating 0.5% (the 5 μL) getting roughly 50 colonies was the goal (which corresponds to 10,000 from the total mixture, which, in a pool of 14 plasmids means each plasmid is represented by 700 colonies.) The rest of the transformation was grown up in 25 mL LB-Carb at 30°C overnight on a roller drum. We miniprepped plasmid from 1.5 mL of these cells (Zyppy™ Plasmid Miniprep Kit) and eluted in 30 μL.

Throughout sequencing all DNA concentration measurements were used using the Qubit 4 Fluorometer with the Qubit™ dsDNA BR Assay Kit (Life Technologies, cat # Q32853). After quantification we used 200 ng of plasmid DNA to make each of the libraries, as per the manufacturer's instructions using the NEBNext® Ultra™ II FS DNA Library Prep Kit for Illumina (New England Biolabs, cat. #7805L), with 150-350bp size selection. We sued NEBNext® Multiplex Oligos for Illumina® Index Primers Sets 1 and 2 (NEB cat# E7335 and E7500), used 4 cycles at PCR step, used Ampure beads for size selection (Beckman Coulter, cat. #63880), with the magnet (DYNAL™ DynaMag™ Dynabeads™ DynaMag -2 Magnet, Invitrogen cat. #12321D) for 1.5 mL microcentrifuge tubes, and at the end the average DNA fragment size of the libraries were validated using the Agilent BioAnalyzer High Sensitivity DNA Kit (Agilent Cat#5067-4626). For each sample, we obtained roughly 20 million 36-bp paired end reads.

Preparation of total RNA and the sample for RNA sequencing :

To prepare the total RNA for sequencing, 500 µL of the clarified lysate prepared was treated with 500 µL TRIzol™ Reagent (ThermoFisher, cat. #15596026) with the addistion of 0.5 µL GenElute™-LPA (Sigma cat. #56575). This was often stored at -80°C until use after mixing by inversion and incubation for 10 minutes at room temperature. When ready to prepare the RNA and the RNA sequencing libraries, precautions were taken to protect the samples, such as use of filter tips as treat all surfaces with RNaseZap® RNase Decontamination Solution (ThermoFisher cat# AM9780). After incubating the RNA in Trizol, we added 100 µL chloroform to each sample, shook well for 15 seconds, incubated the sample at room temperature for 3 minute,s and centrifuged it at 12,000 x g for 15 minutes at 4°C. At this point, the RNA separated into top layer, which was the clear aqueous phase, which was carefully removed and put in another clear (uncolored) 1.5 mL microcentrifuge tube. We then added 250 µL 100% isopropanol, let it incubate at room temperature for 10 minutes, centrifuges it at 12,000 x g, 10 minutes, at 4°C and could then see a small white pellet on bottom of tube, which contained the RNA. Carefully not to lose the pellet, we removed supernatant, washed it with 500 µL 75% EtOH, vortexed, centrifuged at 7,500 x g, 5 minutes, at 4°C and discarded (pipette away) the supernatant. We then let the pellet air dry for 10 minutes, resuspend in 10 µL UltraPure DNase/RNase-Free Distilled Water (ThermoFisher, cat# 10977023) and pipetted up and down. It was then incubated at 55°C for 15 minutes. The RNA concentration was then quantified, which throughout these methods was always done using Qubit 4 Fluorometer with the Qubit™ RNA HS Assay Kit (Life Technologies, cat # Q32855). The RNA was used for preparing the sequencing library or stored at -80°C.

We used roughly 60 ng of purified RNA per library preparation, and did so by following NEBNext® Ultra™ II RNA Library Prep Kit for Illumina® (Chapter 4, NEB cat #E7770). We always ended up using 6 cycles at the PCR step. We did not enrich or deplete the total RNA sample in any way (but in the future rRNA being removed might be favorable). Many of the same reagents were used with the RNA sequencing kit as described in the plasmid DNA sequencing protocol above (Ampure beads, magnet, primers, BioAnalyzer). For each sample, we obtained roughly 800 million 36-bp paired end reads (needed this high amount to get enough coverage of the L1 RNA in the total RNA pool).

Preparation of IP'd RNA and the sample for RNA sequencing :

To prepare the IP'd RNA for sequencing, 500 µL of the clarified lysate prepared was incubated with 10 µL of anti-ORF1 antibody (Milliprep, prod. number MABC1152) - conjugated Dynabeads (30 mg/mL; used Dynabeads Antibody Coupling kit, Life Technologies, prod. number 14311D). The IP was incubated at 4°C on a tube rotator for constant mixing for 1 hour. The beads were then washed 3 times with the same Extraction Buffer, and after thoroughly removing the buffer from the final wash step, the RNA was eluted off the beads by adding 500 µL TRIzol™ Reagent. The protocol through preparing the sequencing library is the same from there as preparing the total RNA samples. For each sample, we obtained roughly 80 million 36-bp paired end reads.

Sequencing and customized alignment strategy of DNA and RNA reads for RNP formation analyses

All sequencing was done using 36bp paired end reads on the NextSeq 500 machine. The libraries were mixed in Illumina Buffer RSB (in the ratios needed to obtain the desired number of reads) and loaded onto Illumina flow cells (NextSeq 500/550 Hi Output KT v2.5 (75 Cycle), Illumina cat#20024906). The data were managed and de-multiplexed by Dr. Matt Maurano,

For analysis, we designed a custom series of L1 reference sequences corresponding to each L1 trialanine variant. The references were designed for each variant : (1) with the mutant sequence (9bp) located at the center of a 75bp sequence (with 35 bp of WT L1 on either side) and (2) the exact same sequence that was fully WT. The 36 bp reads only needed 1 bp of overlap with the mutant sequence to map well.

Illumina Nextseq 500 (36 nucleotide paired-end reads) raw fastq files were aligned to custom references (two references per variant in the sample) using the Burrows-Wheeler Alignment (BWA) mem algorithm 84. Alignments were performed using high-stringency cut-offs, with options corresponding to a severe mismatch penalty (-B 38) and a strict seed length (-k 35). The seed length (35 nucleotides) was chosen to ensure that the short 36 nucleotide read length spans at least 1 nucleotide of the 3X alanine mutation region. The output sam file was converted to a compressed bam file using samtools 85. Mutation counts were obtained by counting the max number of reads aligning to the trialanine mutation (or WT 9bp tag) region using samtools depth with a -max option, and a custom awk script (written by David Truong and stored in his database). To obtain normalized IP'd RNA of each mutant L1 in the pool, each trialanine mutation within a pool was normalized to the tagged-WT control within each pool. Ratios for RNA expression and RNP pull-down were each re-normalized by the ratio of DNA plasmids for each mutant within a pool.

Quantification of ORF1p cellular localization

HeLa M2 cells were seeded and transfected into a 96-well tissue culture plate (Sigma, prod. number CLS3799) as described previously in the retrotransposition methods. The cells were treated with 1 μg/ml puromycin 24 hours post-transfection. 48 hours post-transfection the cells were split 1:5 into a glass-bottom 96-well plate (Brooks, prod. number MGB096-1-2-LG-L) with 1 μg/ml doxycycline treatment to induce expression of L1 in addition to continued 1 μg/ml puromycin treatment. 72 hours post-induction the cells were prefixed by adding 10% formalin solution (Sigma, prod. number HT5012) directly to the culture media to a final concentration of 1%. After 10 min at room temperature the media/formaldehyde mixture was discarded and cells were fixed for 20 min at room temperature with 4% formalin in DPBS (Fisher, prod. number 14190-250). Cells were then washed three times in DPBS supplemented with 10 mM glycine and 0.02% sodium azide, pH 7.4 and once in DPBS. The DPBS was then removed and the cells were permeabilized with DPBS with 0.2% Triton X-100 for 20 minutes at room temperature. The permeabilization solution was then removed and the cells were washed five times with DPBS supplemented with 10 mM glycine and 0.02% sodium azide, pH 7.4. The cells were then incubated for at least 1 hour at room temperature in LI-COR Odyssey blocking buffer (LI-COR, prod. number 927-40000). Following blocking, cells were incubated overnight at 4°C with primary antibodies for human ORF1 antibody and for fibrillarin, the nucleolar marker (Milliprep, prod. number MABC1152 and: Abcam, prod. number ab5821, respectively) diluted in LI-COR Odyssey blocking buffer (final antibody concentrations of 1.25 μg /mL and 1 μg /mL, respectively). The next day cells were washed five times in DPBS with 0.1% Triton X-100 and then incubated in the dark with secondary antibodies (the goat anti-mouse IgG H&L - Cy3 : Abcam, prod. number ab97035 and the goat anti-rabbit IgG H&L- Cy5, Abcam, prod. number ab6564; diluted to final concentration of 0.5 μg/mL and 1 μg/mL , respectively) at room temperature for 1-2 hours. The cells were then washed five times in DPBS with 0.1% Triton X-100 and three times in DPBS. The cells were incubated in the dark in Hoechst 33342 solution (Thermo, prod. number 62249) diluted to 400 ng/mL in DPBS for 30 min at room temperature. The Hoechst solution was then removed and DPBS was added to the cells.

Images were obtained using an Andor Yokogawa CSU-X confocal spinning disk on a Nikon TI Eclipse microscope and fluorescence was recorded with an sCMOS Prime 95B camera (Photometrics) with a 100x objective (pixel size: 0.11 μm). Images were acquired using Nikon Elements software and analyzed using ImageJ/Fiji. Nucleolar phenotype was qualitatively evaluated by normalizing a given cell's nucleolar ORF1p intensity to its cytoplasmic ORF1p intensity and comparing it to the same ratio in cells transfected with the wild-type construct. A "dim" phenotype corresponded to cells with a nucleolar-to-cytoplasmic ORF1p intensity ratio of between 0.8 and 1.2 with clear labeling of the nucleoli, while a "bright" phenotype described cells with a ratio greater than 1.2. Nucleoli were identified by DAPI and were confirmed by fibrillarin immunofluorescence in a subset of experiments. The frequency of the nucleolar phenotype was evaluated over at least 5 fields of view per construct per experiment.

Alignments and categorization of used for analysis conservation in ORF2p

Reliable ORF2p sequences were provided by Stephane Boissinot, the GenBank IDs of which are listed in Table 3.9. WT human L1-rp was compared to the indicated mammalian sequences only and then, more interestingly, all 55 of the sequences listed were run through multiple sequence alignment analysis followed by measurements of percent identity using Geneious (more specifically Geneious® 11.1.2; Build 2018-03-01 15:52; Java Version 1.8.0_162-b12 (64 bit): Restricted R11 license). The program produced the % identity score at each residue. Since we are working with three residue window, we chose to use the identity % score value that corresponded to the residue with the highest identity score each trialanine variant. We binned the identify score quantities into four bins, spanning from lowest to highest : 0-29%, 30-69%, 70-99%, and 100%. We then compared these categories to the three bins of retrotransposition efficiency explained in the text (no retrotransposition, reduced retrotransposition, and WT levels of retrotransposition). Then, the status of each variant by each of these two measures was analyzed.

Page intentionally left blank.

References

1. Kazazian, H. H. Mobile Elements: Drivers of Genome Evolution. *Science* (2004). doi:10.1126/science.1089670

2. Faulkner, G. J. & Garcia-Perez, J. L. L1 Mosaicism in Mammals: Extent, Effects, and Evolution. *Trends Genet.* **33,** 802–816 (2017).

3. Huang, C. R. L., Burns, K. H. & Boeke, J. D. Active Transposition in Genomes. *Annu. Rev. Genet.* **46,** 651–675 (2012).

4. Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* (2001). doi:10.1038/35057062

5. Beck, C. R. *et al.* LINE-1 retrotransposition activity in human genomes. *Cell* (2010). doi:10.1016/j.cell.2010.05.021

6. Szak, S. T. *et al.* Molecular archeology of L1 insertions in the human genome. *Genome Biol.* (2002). doi:10.1186/gb-2002-3-10-research0052

7. Ostertag, E. M. & Kazazian, H. H. Biology of Mammalian L1 Retrotransposons. *Annu. Rev. Genet.* **35,** 501–538 (2001).

8. Brouha, B. *et al.* Hot L1s account for the bulk of retrotransposition in the human population. *Proc. Natl. Acad. Sci.* (2003). doi:10.1073/pnas.0831042100

9. Muotri, A. R. *et al.* Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. *Nature* (2005). doi:10.1038/nature03663

10. Carreira, P. E., Richardson, S. R. & Faulkner, G. J. L1 retrotransposons, cancer stem cells and oncogenesis. *FEBS J.* **281,** 63–73 (2014).

11. Hancks, D. C. & Kazazian, H. H. Roles for retrotransposon insertions in human disease. *Mobile DNA* (2016). doi:10.1186/s13100-016-0065-9

12. Ardeljan, D., Taylor, M. S., Ting, D. T. & Burns, K. H. The human long interspersed element-1 retrotransposon: An emerging biomarker of Neoplasia. *Clin. Chem.* **63,** 816–822 (2017).

13. Burns, K. H. Transposable elements in cancer. *Nat. Rev. Cancer* **17,** 415–424 (2017).

14. Rodić, N. *et al.* Long interspersed element-1 protein expression is a hallmark of many human cancers. *Am. J. Pathol.* (2014). doi:10.1016/j.ajpath.2014.01.007

15. Nguyen, T. H. M. *et al.* L1 Retrotransposon Heterogeneity in Ovarian Tumor Cell Evolution. *Cell Rep.* (2018). doi:10.1016/j.celrep.2018.05.090

16. Doucet-O'Hare, T. T. *et al.* LINE-1 expression and retrotransposition in Barrett's esophagus and esophageal carcinoma. *Proc. Natl. Acad. Sci.* (2015). doi:10.1073/pnas.1502474112

17. Lee, E. *et al.* Landscape of somatic retrotransposition in human cancers. *Science (80-.).* (2012). doi:10.1126/science.1222077

18. Gorbunova, V., Boeke, J. D., Helfand, S. L. & Sedivy, J. M. Sleeping dogs of the genome. *Science (80-.).* **346,** 1187–1188 (2014).

19. Swergold, G. D. Identification, characterization, and cell specificity of a human LINE-1 promoter. *Mol. Cell. Biol.* (1990). doi:10.1128/MCB.10.12.6718

20. Speek, M. Antisense Promoter of Human L1 Retrotransposon Drives Transcription of Adjacent Cellular Genes. *Mol. Cell. Biol.* (2001). doi:10.1128/MCB.21.6.1973-1985.2001

21. Doucet, A. J., Wilusz, J. E., Miyoshi, T., Liu, Y. & Moran, J. V. A 3' Poly(A) Tract Is Required for LINE-1 Retrotransposition. *Mol. Cell* (2015). doi:10.1016/j.molcel.2015.10.012

22. Dombroski, B. A., Mathias, S. L., Nanthakumar, E., Scott, A. F. & Kazazian, H. H. Isolation of an active human transposable element. *Science (80-.).* (1991). doi:10.1126/science.1662412

23. Burns, K. H. & Boeke, J. D. Human transposon tectonics. *Cell* **149,** 740–752 (2012).
24. Denli, A. M. *et al.* Primate-Specific ORF0 Contributes to Retrotransposon-Mediated Diversity. *Cell* (2015). doi:10.1016/j.cell.2015.09.025

25. Kolosha, V. O. & Martin, S. L. In vitro properties of the first ORF protein from mouse LINE-1 support its role in ribonucleoprotein particle formation during retrotransposition. *Proc. Natl. Acad. Sci. U. S. A.* (1997). doi:10.1073/pnas.94.19.10155

26. Martin, S. L. & Bushman, F. D. Nucleic Acid Chaperone Activity of the ORF1 Protein from the Mouse LINE-1 Retrotransposon. *Mol. Cell. Biol.* (2001). doi:10.1128/MCB.21.2.467-475.2001

27. Feng, Q., Moran, J. V., Kazazian, H. H. & Boeke, J. D. Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. *Cell* (1996). doi:10.1016/S0092-8674(00)81997-2

28. Mathias, S. L., Scott, A. F., Kazazian, H. H., Boeke, J. D. & Gabriel, A. Reverse transcriptase encoded by a human transposable element. *Science (80-.).* (1991). doi:10.1126/science.1722352

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

30. Wei, W. *et al.* Human L1 Retrotransposition: cis Preference versus trans Complementation. *Mol. Cell. Biol.* (2001). doi:10.1128/MCB.21.4.1429-1439.2001

31. Kulpa, D. A. & Moran, J. V. Cis-preferential LINE-1 reverse transcriptase activity in ribonucleoprotein particles. *Nat. Struct. Mol. Biol.* (2006). doi:10.1038/nsmb1107

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

33. Martin, S. L. Ribonucleoprotein Particles with LINE-1 RNA in Mouse Embryonal Carcinoma Cells. *Mol. Cell. Biol.* **11,** 4804–4807 (1991).

34. Hohjohl, H. & F. Singer, M. Cytoplasmic ribonucleoprotein complexes containing human LINE-1 protein and RNA. *EMBO J.* (1996). doi:10.1002/j.1460-2075.1996.tb00395.x

35. Kulpa, D. A. & Moran, J. V. Ribonucleoprotein particle formation is necessary but not sufficient for LINE-1 retrotransposition. *Hum. Mol. Genet.* (2005). doi:10.1093/hmg/ddi354

36. Doucet, A. J. *et al.* Characterization of LINE-1 ribonucleoprotein particles. *PLoS Genet.* (2010). doi:10.1371/journal.pgen.1001150

37. Taylor, M. S. *et al.* Affinity proteomics reveals human host factors implicated in discrete stages of LINE-1 retrotransposition. *Cell* **155,** 1034–1048 (2013).

38. Taylor, M. S. *et al.* Dissection of affinity captured LINE-1 macromolecular complexes. *Elife* (2018). doi:10.7554/eLife.30094

39. Cost, G. J., Feng, Q., Jacquier, A. & Boeke, J. D. Human L1 element target-primed reverse transcription in vitro. *EMBO J.* (2002). doi:10.1093/emboj/cdf592

40. Luan, D. D., Korman, M. H., Jakubczak, J. L. & Eickbush, T. H. Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: A mechanism for non-LTR retrotransposition. *Cell* (1993). doi:10.1016/0092-8674(93)90078-5

41. Mita, P. *et al.* LINE-1 protein localization and functional dynamics during the cell cycle. *Elife* (2018). doi:10.7554/eLife.30058

42. Niewiadomska, A. M. *et al.* Differential Inhibition of Long Interspersed Element 1 by APOBEC3 Does Not Correlate with High-Molecular-Mass-Complex Formation or P-Body Association. *J. Virol.* (2007). doi:10.1128/JVI.02800-06

43. Liu, N. *et al.* Selective silencing of euchromatic L1s revealed by genome-wide screens for L1 regulators. *Nature* **553,** 228–232 (2018).

44. Suzuki, J. *et al.* Genetic evidence that the non-homologous end-joining repair pathway is involved in LINE retrotransposition. *PLoS Genet.* (2009). doi:10.1371/journal.pgen.1000461

45. Peddigari, S., Li, P. W. L., Rabe, J. L. & Martin, S. L. HnRNPL and nucleolin bind LINE-1 RNA and function as host factors to modulate retrotransposition. *Nucleic Acids Res.* (2013). doi:10.1093/nar/gks1075

46. Dai, L., Taylor, M. S., O'Donnell, K. A. & Boeke, J. D. Poly(A) Binding Protein C1 Is Essential for Efficient L1 Retrotransposition and Affects L1 RNP Formation. *Mol. Cell. Biol.* (2012). doi:10.1128/MCB.06785-11

47. Pizarro, J. G. & Cristofari, G. Post-Transcriptional Control of LINE-1 Retrotransposition by Cellular Host Factors in Somatic Cells. *Front. Cell Dev. Biol.* (2016). doi:10.3389/fcell.2016.00014

48. Goodier, J. L., Cheung, L. E. & Kazazian, H. H. Mapping the LINE1 ORF1 protein interactome reveals associated inhibitors of human retrotransposition. *Nucleic Acids Res.* (2013). doi:10.1093/nar/gkt512

49. Beauregard, A., Curcio, M. J. & Belfort, M. The Take and Give Between Retrotransposable Elements and their Hosts. *Annu. Rev. Genet.* (2008). doi:10.1146/annurev.genet.42.110807.091549

50. Goodier, J. L., Cheung, L. E. & Kazazian, H. H. MOV10 RNA Helicase Is a Potent Inhibitor of Retrotransposition in Cells. *PLoS Genet.* (2012). doi:10.1371/journal.pgen.1002941

51. Arjan-Odedra, S., Swanson, C. M., Sherer, N. M., Wolinsky, S. M. & Malim, M. H. Endogenous MOV10 inhibits the retrotransposition of endogenous retroelements but not the replication of exogenous retroviruses. *Retrovirology* (2012). doi:10.1186/1742-4690-9-53

52. Han, J. S. & Boeke, J. D. A highly active synthetic mammalian retrotransposon. *Nature* (2004). doi:10.1038/nature02535

53. An, W. *et al.* Characterization of a synthetic human LINE-1 retrotransposon ORFeus-Hs. *Mob. DNA* **2,** 2 (2011).

54. Ostertag, E. M. *et al.* A mouse model of human L1 retrotransposition. *Nat. Genet.* (2002). doi:10.1038/ng1022

55. An, W. *et al.* Active retrotransposition by a synthetic L1 element in mice. *PNAS* (2006). doi:10.1073/pnas.0605300103

56. Donnell, K. A. O., An, W., Schrum, C. T., Wheelan, S. J. & Boeke, J. D. Controlled insertional mutagenesis using a LINE-1 (ORFeus) gene-trap mouse model. *Proc. Natl. Acad. Sci.* (2013). doi:10.1073/pnas.1302504110

57. Kano, H. *et al.* L1 retrotransposition occurs mainly in embryogenesis and creates somatic mosaicism. *Genes Dev.* (2009). doi:10.1101/gad.1803909

58. Richardson, S. R. *et al.* Heritable L1 retrotransposition in the mouse primordial germline and early embryo. *Genome Res.* (2017). doi:10.1101/gr.219022.116

59. LaCava, J., Jiang, H. & Rout, M. P. Protein Complex Affinity Capture from Cryomilled Mammalian Cells. *J. Vis. Exp.* 1–11 (2016). doi:10.3791/54518

60. Rothbauer, U. *et al.* A Versatile Nanotrap for Biochemical and Functional Studies with Fluorescent Fusion Proteins. *Mol. Cell. Proteomics* (2008). doi:10.1074/mcp.M700342-MCP200

61. Khazina, E. *et al.* Trimeric structure and flexibility of the L1ORF1 protein in human L1 retrotransposition. *Nat. Struct. Mol. Biol.* **18,** 1006–1014 (2011).

62. Khazina, E. & Weichenrieder, O. Human LINE-1 retrotransposition requires a metastable coiled coil and a positively charged N-terminus in L1ORF1p. *Elife* **7,** 1–29 (2018).

63. Christian, C. M., Deharo, D., Kines, K. J., Sokolowski, M. & Belancio, V. P. Identification of L1 ORF2p sequence important to retrotransposition using Bipartile Alu retrotransposition (BAR). *Nucleic Acids Res.* **44,** 4818–4834 (2016).

64. Weichenrieder, O., Repanas, K. & Perrakis, A. Crystal structure of the targeting endonuclease of the human LINE-1 retrotransposon. *Structure* (2004). doi:10.1016/j.str.2004.04.011

65. Januszyk, K. *et al.* Identification and solution structure of a highly conserved C-terminal domain within ORF1p required for retrotransposition of long interspersed nuclear element-1. *J. Biol. Chem.* **282,** 24893–24904 (2007).

66. Khazina, E. & Weichenrieder, O. Non-LTR retrotransposons encode noncanonical RRM domains in their first open reading frame. *Proc. Natl. Acad. Sci.* (2009). doi:10.1073/pnas.0809964106

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

68. Fanning, T. & Singer, M. The line-1 DNA sequences in four mammalian orders predict proteins that conserve homologies to retrovirus proteins. *Nucleic Acids Res.* (1987). doi:10.1093/nar/15.5.2251

69. Daugherty, M. D. & Malik, H. S. Rules of Engagement: Molecular Insights from Host-Virus Arms Races. *Annu. Rev. Genet.* **46,** 677–700 (2012).

70. Caudron-Herger, M. *et al.* Alu element-containing RNAs maintain nucleolar structure and function. *EMBO J.* (2015). doi:10.15252/embj.201591458

71. Ahl, V., Keller, H., Schmidt, S. & Weichenrieder, O. Retrotransposition and Crystal Structure of an Alu RNP in the Ribosome-Stalling Conformation. *Mol. Cell* (2015). doi:10.1016/j.molcel.2015.10.003

72. Bernardi, R. *et al.* PML regulates p53 stability by sequestering Mdm2 to the nucleolus. *Nat. Cell Biol.* (2004). doi:10.1038/ncb1147

73. Condemine, W., Takahashi, Y., Le Bras, M. & de The, H. A nucleolar targeting signal in PML-I addresses PML to nucleolar caps in stressed or senescent cells. *J. Cell Sci.* (2007). doi:10.1242/jcs.007492

74. Dutrieux, J. *et al.* PML/TRIM19-Dependent Inhibition of Retroviral Reverse-Transcription by Daxx. *PLoS Pathog.* **11,** 1–22 (2015).

75. Watanabe, K. *et al.* Rad18 guides polη to replication stalling sites through physical interaction and PCNA monoubiquitination. *EMBO J.* (2004). doi:10.1038/sj.emboj.7600383

76. Ariumi, Y. *et al.* DNA repair protein Rad18 restricts LINE-1 mobility. *Sci. Rep.* (2018). doi:10.1038/s41598-018-34288-9

77. Inagaki, A. *et al.* Dynamic localization of human RAD18 during the cell cycle and a functional connection with DNA double-strand break repair. *DNA Repair (Amst).* (2009). doi:10.1016/j.dnarep.2008.10.008

78. Ostertag, E. M., Prak, E. T., DeBerardinis, R. J., Moran, J. V & Kazazian, H. H. Determination of L1 retrotransposition kinetics in cultured cells. *Nucleic Acids Res.* (2000). doi:gkd248 [pii]

79. Szent-Gyorgyi, C. *et al.* Fluorogen-activating single-chain antibodies for imaging cell surface proteins. *Nat. Biotechnol.* (2008). doi:10.1038/nbt1368

80. Wang, Y. *et al.* Fluorogen activating protein-affibody probes: Modular, no-wash measurement of epidermal growth factor receptors. *Bioconjug. Chem.* (2015). doi:10.1021/bc500525b

81. Richardson, S. M., Wheelan, S. J., Yarrington, R. M. & Boeke, J. D. GeneDesign: Rapid, automated design of multikilobase synthetic genes. *Genome Res.* (2006). doi:10.1101/gr.4431306

82. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* (2009). doi:10.1038/nmeth.1318

83. Hampf, M. & Gossen, M. Promoter Crosstalk Effects on Gene Expression. *J. Mol. Biol.* (2007). doi:10.1016/j.jmb.2006.10.009

84. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* (2010). doi:10.1093/bioinformatics/btp698

85. Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* (2011). doi:10.1093/bioinformatics/btr509

Page intentionally left blank.

Curriculum Vitae

CV spans the next 5 pages.

EMILY MARIE ADNEY

JOHNS HOPKINS SCHOOL OF MEDICINE • BALTIMORE, MD
NEW YORK UNIVERSITY SCHOOL OF MEDICINE • NEW YORK, NY
929.307.5329 • EADNEY@GMAIL.COM

EDUCATION

Johns Hopkins University School of Medicine

- **Ph.D. in Human Genetics**
- **Graduate student in the Institute of Genetic Medicine Human Genetics Pre-Doctoral Training Program**

  July 2012 – December 2018

Johns Hopkins University Krieger School of Arts & Sciences

- **M.S. in Molecular Biophysics**

  November 2011

Occidental College

- **B.A. in Molecular Biology**

  May 2009

RESEARCH & PROFESSIONAL EXPERIENCE

Graduate Research

July 2012– December 2018

Human Genetics Pre-Doctoral Training Program, Johns Hopkins University
Thesis Advisor : Jef Boeke, Ph.D.

- Used high-throughput approaches to both create a comprehensive ordered mutagenic library of the L1 mobile element and test the activity of each mutant construct using in-cell, biochemical, and DNA sequencing-based assays
- Completed extensive screens of how overexpression and knock down of hundreds of genes affects the activity of the L1 mobile element
- Developed new transgenic mouse models and monoclonal antibodies for proteomic analysis of active L1 mobile elements

July 2009 – Nov 2011

Program in Molecular Biophysics, Johns Hopkins University
Thesis Advisor : Juliette Lecomte, Ph.D.

- Used nuclear magnetic resonance and other spectroscopy methods to analyze the 3D structure of variants of cyanobacterial hemoglobins
- Customized protein purification protocols for hemoglobins and obtained high quality samples for subsequent biochemistry

Graduate Internship

April 2015 – November 2015

NYU Sackler Office of Industrial Liaison Biotechnology Internship

- Completed coursework provided by members of Technology Transfer Office
- Prepared a formal business plan for a new company based on a provided patent

Undergraduate Research

Summer 2009

The Mayo Clinic Summer Undergraduate Research Fellowship

Advisor : Zhiguo Zhang, Ph.D.

Summer 2008

Children's Hospital of Oakland Research Institute Summer Student Research Program Internship

Advisor : Dario Boffelli, M.D.

PUBLICATIONS & POSTERS

Publications :

- Comprehensive scanning mutagenesis of a human retrotransposon LINE-1 identifies motifs essential for function. **Emily M. Adney**, David M. Truong, Srinjoy Sil, Matthias T. Ochmann, Paolo Mita, Xuya Wang, Zoltán Ivics, David Fenyö, Liam Holt, and Jef D. Boeke. (*In preparation, submission 2018.)*

- Prostate-specific loss of UXT promotes cancer progression. Yu Wang , Eric Schafler, Phillip Thomas, Susan Ha, Gregory David, **Emily M. Adney** , Michael Garabedian, Peng Lee, Susan Logan, *Oncotarget, 2018.*

- LINE-1 elements are derepressed in senescent cells and elicit a chronic Type-I Interferon response. Marco De Cecco, Takahiro Ito, Amy E. Elias, Nicholas J. Skvir, Steven W. Criscione, Alberto Caligiana, Greta Brocculi, **Emily M. Adney**, Jef D. Boeke, Jayakrishna Ambati, Matthew Simon, Andrei Seluanov, Vera Gorbunova, Eline Slagboom, Stephen L. Helfand, Nicola Neretti, John M. Sedivy, *Nature, 2018.*

- Dissection of affinity captured LINE-1 macromolecular complexes. Martin S Taylor, Ilya Altukhov, Kelly R Molloy, Paolo Mita , Hua Jiang , **Emily M. Adney**, Aleksandra Wudzinska, Sana Badri, Dmitry Ischenko, George Eng, Kathleen H Burns, David Fenyö, Brian T Chait, Dmitry Alexeev, Michael P Rout, Jef D Boeke, John LaCava, *eLife,* Jan 2018.

- Affinity proteomics reveals human host factors implicated in discrete stages of LINE-1 retrotransposition. Martin S. Taylor, John LaCava, Paolo Mita, Kelly R. Molloy, Lisa Huang Cheng, Donghui Li, **Emily M. Adney**, Hua Jiang, Kathleen H. Burns, Brian T. Chait, Jef D. Boeke , Lixin Dai, *Cell*, Nov 2013.

- Chemical reactivity of Synechococcus sp. PCC 7002 and Synechocystis sp. PCC 6803 hemoglobins: covalent heme attachment and bishistidine coordination. Henry J. Nothnagel, Matthew R. Preimesberger, Matthew P. Pond, Benjamin Y. Winer, **Emily M. Adney**, and Juliette T. J. Lecomte, *J Biol Inorg Chem*, Apr 2011.

Posters :

- 2017 FASEB Conference : Mobile DNA in Mammalian Genomes
  (Big Sky, Montana)
  "An Ordered Mutagenic Library of Human L1" **Emily M. Adney**, Matthias T. Ochmann, Xuya Wang, David Fenyö, Zoltán Ivics, and Jef D. Boeke

- 2015 FASEB Conference : Mobile DNA in Mammalian Genomes
  (Palm Beach, Florida)
  "Insight into How the Life Cycle of LINE1 Ribonucleoprotein Complexes is Impacted by High-Confidence Candidate Protein Interactors : a Knock Down Screen" **Emily M. Adney**, Donghui Li and Jef D. Boeke

- 2014 Keystone Conference : Mobile Genetic Elements and Genome Evolution
  (Sante Fe, New Mexico)
  "How the Life Cycle of Active LINE1 Ribonucleoprotein Complexes is Impacted by Overexpression of a List of High-Confidence Candidate Protein Interactors" **Emily M. Adney**, Martin S. Taylor, Lixin Dai and Jef D. Boeke

- 2011 Biophysical Society 55th Annual Meeting
  (Baltimore, Maryland)
  "Chemical reactivity of Synechocystis sp. PCC 6803 hemoglobins: covalent heme attachment and bishistidine coordination" Matthew R. Preimesberger, **Emily M. Adney**, and Juliette T. J. Le.comte

OUTREACH & TEACHING EXPERIENCE

Fall 2017 – Present	**Founder and Organizer of ISG Papers and Tea** Institute of Systems Genetics, NYU Langone Health

- Started an institute-wide journal club that meets twice a month
- Organize the schedule, pick up food, and recruit presenters

Winter 2017 – Present

Contributor to Science Sketches
San Francisco, CA and New York, NY

- Participate in this effort to make two-minute videos that teach various topics in science
- Wrote scientific narratives and helped with the artistic representation for several Science Sketches

Fall 2014 - Present

Training of Lab Members
Boeke Lab, NYU Langone Health

- Srinjoy Sil – graduate student
- Matthias Ochmann – graduate student

Spring 2010 – Fall 2011

Training of Lab Members
Lecomte Lab, Johns Hopkins University

- Selena Rice – graduate student
- Belinda Wenke – technician

April 2010

Teacher at the USA Science and Engineering Festival
Washington, DC

- Provided lessons on nuclear magnetic resonance and protein structure to the general public

Fall 2008 - Spring 2009

Organic Chemistry Tutor
Occidental College, the Academic Mastery Program

- Designed workshops to compliment course work and helped students prepare for exams

Spring 2006 - Spring 2008

Lead Science and Math Tutor
Occidental College, Gear Up Tutoring Program

- Mentored middle school students in all subjects

HONORS & AWARDS

- The Mayo Clinic Summer Undergraduate Research Fellowship (2009)
- The Maes Travel Fellowship : Occidental College, Languages and Literature Department (2009)
- Federal National Science and Mathematics Access to Retain Talent (SMART) Grant (2007-2009)
- Minnie K. and William Carter, Jr. Endowed Scholarship for achievement in Biology (2006-2009)
- Federal Pell Grant (2005-2009)

- Occidental College Scholarship (2005-2009)

CODING EXPERIENCE

- **Johns Hopkins University Krieger School of Arts & Sciences**
    - Molecular Biophysics Program's Python Bootcamp
        - 2009 Taught by Gregory Bowman, Ph.D.
- **Johns Hopkins University School of Medicine**
    - Introduction to Python Workshop
        - 2013 Taught by Sarah Wheelan, Ph.D.

REFERENCES

- **Jef D. Boeke, Ph.D.**
  Director of the Institute of Systems Genetics, NYU School of Medicine
    - Relationship : Ph.D. thesis advisor
    - Jef.Boeke@nyumc.org
    - 646-501-0504

- **Kathleen H. Burns, M.D., Ph.D.**
  Professor of Pathology, Johns Hopkins School of Medicine
    - Relationship : long-term collaborator, mentor, and thesis committee member
    - KBurns@jhmi.edu
    - 410-502-7214

- **David Fenyö, Ph.D.**
  Professor of Biochemistry and Molecular Pharmacology, Institute for Systems Genetics, NYU School of Medicine
    - Relationship : long-term collaborator and mentor
    - David.Fenyo@nyumc.org
    - 212-263-2216

Page intentionally left blank.