Identification of L1 ORF2p sequence important to retrotransposition using Bipartile Alu retrotransposition (BAR)

Claiborne M. Christian, Dawn deHaro, Kristine J. Kines, Mark Sokolowski and Victoria P. Belancio[*]

Structural and Cellular Biology, Tulane University School of Medicine, Tulane Cancer Center, Tulane Center for Aging, New Orleans, LA 70130, USA

ABSTRACT

Long Interspersed Element 1 (LINE-1 or L1) is capable of causing genomic instability through the activity of the L1 ORF2 protein (ORF2p). This protein contains endonuclease (EN) and reverse transcriptase (RT) domains that are necessary for the retrotransposition of L1 and the Short Interspersed Element (SINE) Alu. The functional importance of approximately 50% of the ORF2p molecule remains unknown, but some of these sequences could play a role in retrotransposition, or be necessary for the enzymatic activities of the EN and/or RT domains. Conventional approaches using the full-length, contiguous ORF2p make it difficult to study the involvement of these unannotated sequences in the function of L1 ORF2p. Our lab has developed a Bipartile Alu Retrotransposition (BAR) assay that relies on separate truncated ORF2p fragments: an EN-containing and an RT-containing fragment. We validated the utility of this method for studying the ORF2p function in retrotransposition by assessing the effect of expression levels and previously characterized mutations on BAR. Using BAR, we identified two pairs of amino acids important for retrotransposition, an FF and a WD. The WD appears to play a role in cDNA synthesis by the ORF2p molecule, despite being outside the canonical RT domain.

INTRODUCTION

Long Interspersed Element 1 (LINE1 or L1) is the only active, autonomous retroelement within the human genome (Figure 1). L1 and the parasitic Short Interspersed Element (SINE) Alu have amplified in the human genome to approximately 500 000 and 1 000 000, respectively, copies through a copy-and-paste mechanism termed target-primed reverse transcription (TPRT). TPRT requires the enzymatic activities of the L1 ORF2 protein (ORF2p). An Apurinic/Apyrimidic Endonuclease (EN) is present at the N-terminus of the ORF2p (Figure 1) (1). This EN domain creates a free 3′-OH that can be utilized by the reverse transcriptase (RT) domain of the ORF2p. RT reverse transcribes the parent L1 mRNA into cDNA that is then integrated into the host genome as a new L1 copy (2). The mechanisms of second strand nicking and second strand synthesis as well as DNA repair following this process remain poorly understood.

Apart from the enzymatic EN and RT domains, there are two other sequence areas within the ORF2p important to retrotransposition. One of these is the Z domain, located adjacent to the EN domain (Figure 1). This domain contains an octapeptide sequence that is well conserved across the ORF2p and ORF2p-equivalents from a variety of species from all taxa of life (3). In the related R2 retroelement, this octapeptide is part of an RNA binding motif (4). Additionally, the Z domain contains a PCNA binding motif that has been shown to be essential for L1 retrotransposition (5). ORF2p also possesses a C-terminal cysteine-rich domain (Cys) that is important for L1 retrotransposition (Figure 1) (6). This domain may have a role in nucleic acid binding (7). The EN, Z, RT and Cys domains are separated by stretches of protein sequence that represent approximately 50% of the molecule but have no known function (Figure 1).

The ORF2p is thought to only function as a single, full-length, contiguous molecule in its capacity to drive retrotransposition. Recent data have shown that fragments of ORF2p retain the catalytic activity of the EN domain in mammalian cells (8). The importance of the sequence C-terminal to the EN domain and N-terminal to the Z domain to retrotransposition has not been previously described. The addition of this sequence and the Z domain to the EN domain decreased toxicity due to the EN-induced DNA damage when compared to the ORF2p EN domain alone

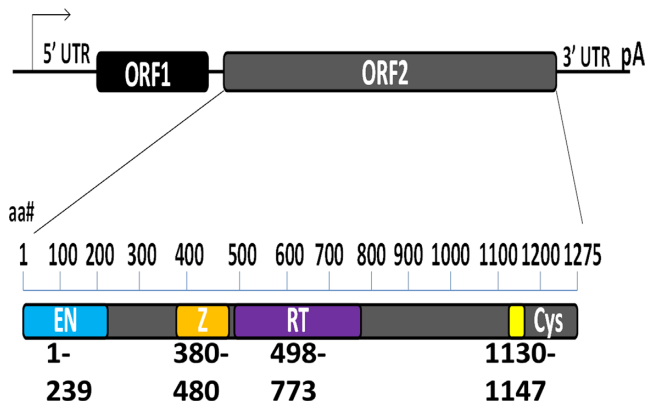[*]To whom correspondence should be addressed. Tel: +1 504 988 4506; Fax: +1 504 988 6684; Email: vperepe@tulane.edu

Figure 1. Long Interspersed Element 1 (LINE-1/L1) and ORF2. A schematic representation of L1, including the transcription start site in the 5′ untranslated region (5′ UTR), Open Reading Frame 1 (ORF1), Open Reading Frame 2 (ORF2), the 3′ untranslated region (3′ UTR), and poly a signal (pA). Inset is a schematic of the protein encoded for by the ORF2, the ORF2 protein (ORF2p) drawn to scale with amino acid (aa) number. The ORF2p consists of an endonuclease domain (EN: blue), Z domain (Z: orange) that includes a PCNA binding sequence, reverse transcriptase domain (RT: purple), and cysteine-rich domain (Cys: yellow) that may be involved in nucleic acid binding. These domains represent ∼50% of the ORF2p sequence and contain both catalytic activities (EN and RT) and noncatalytic functions (Z and Cys) amino acids important for retrotransposition. The rest of the ORF2p (∼50%: gray) has no known function in retrotransposition. Amino acid positions of domains indicated below each domain. L1PA1 ORF2p sequence used.

Figure 2. Fragments of the ORF2p expressed from different plasmids can support retrotransposition of the Short Interspersed Element (SINE) Alu. (**A**) Schematic representation of the ORF2 fragments. The ORF2 is shown for reference. The ENZ490Δ fragment contains the EN and Z domains (amino acids 1 through 490). The Δ348RTCys fragment contains the Z, RT and Cys domains (amino acids 348 through 1275). Constructs were generated using codon optimized ORF2 DNA sequence. (**B**) Western blot analysis of ENZ490Δ and Δ348RTCys protein expression in HeLa cells with custom ORF2p antibodies. Bands of expected molecular weight for each construct denoted by black asterisks. Control lane (C) is total cell lysate from HeLa cells transfected with empty vector. Molecular weight markers indicated at left. GAPDH used as loading control. Bands of smaller molecular weight may represent truncated/processed/degraded ORF2p fragment products. (**C**) Schematic of the Bipartile Alu Retrotransposition (BAR) assay. (**D**) Results of BAR assay in HeLa cells. Neo^r colonies correspond to *de novo* retrotransposition events. Error bars denote standard deviation of $n = 3$ experiments. Representative flasks are shown above corresponding graph bars.

(8). These data suggest that the sequence between the EN and RT domains may affect the function of the EN domain, and by extension retrotransposition. Understanding the contribution of these unannotated regions of the ORF2p to L1 and Alu amplification would significantly expand our knowledge of the ORF2p function during retrotransposition in mammalian cells. It would also open opportunities to identify specific cellular factors that may be influencing retrotransposition *via* this sequence. Using the full-length ORF2p to identify such areas of importance is challenging, as the enzymatic functions of the ORF2p are physically coupled. This coupling makes it difficult to discern whether mutations to the ORF2p outside of the enzymatic domains have an effect on one, both, or neither of their separate enzymatic functions.

To fill this gap in knowledge, we developed an assay that would allow us to elucidate the importance of unannotated sequences to retrotransposition. To do so, we developed a system that we term Bipartile Alu Retrotransposition (BAR). In BAR, expression plasmids containing 5′- or 3′-truncated fragments of human ORF2 DNA sequence (Figure 2A) are supplied *in trans* with the Alu retrotransposition reporter construct (Figure 2B). This allows for assessment of the two N- or C-terminally truncated ORF2p fragments to cooperate in driving Alu retrotransposition, facilitating the interrogation of the importance of unannotated ORF2p sequences to the function of these fragments. BAR has the advantage of allowing for the analysis of ORF2p sequence independent of the L1 ORF1p, as Alu retrotransposition is enhanced by ORF1p, but does not strictly require it (9). This is in contrast to L1 retrotransposition, which does require the ORF1p (6). Our data
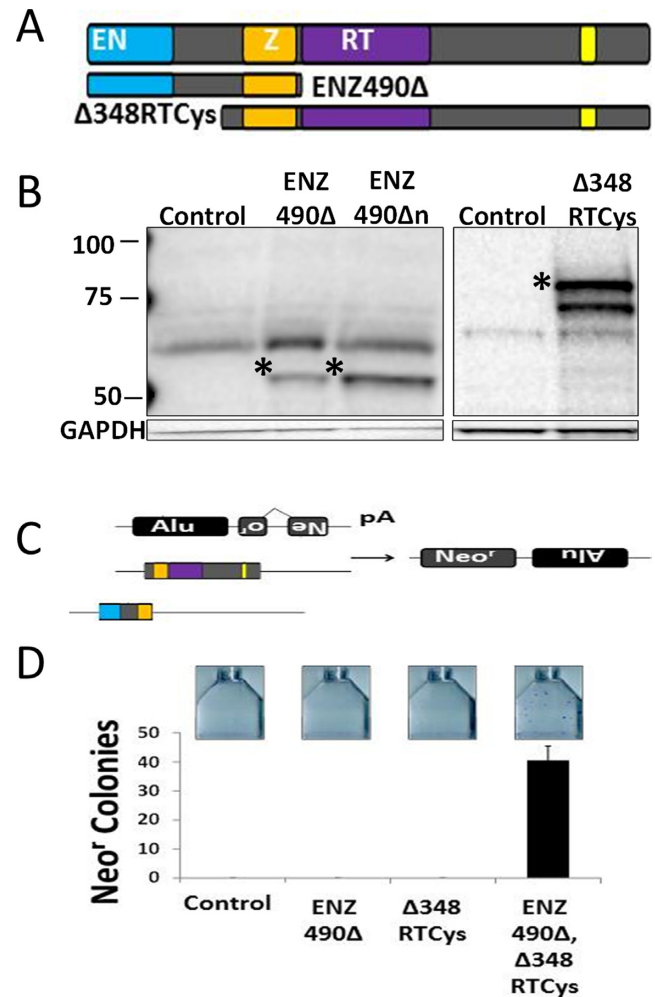
demonstrate that the ORF2p can be split into two, separately expressed protein fragments to drive Alu retrotransposition. We validated the utility of this method for advancing our knowledge of the ORF2p function in retrotransposition by characterizing a previously described PCNA binding domain mutation in BAR (5). We were then able to use BAR to identify novel areas of the ORF2p important for retrotransposition. Our findings demonstrate that the BAR method is useful for the more detailed characterization of

already known as well as newly discovered ORF2p functions in retrotransposition. Specifically, BAR allows us to easily manipulate smaller ORF2p fragments, detect such fragments using a single antibody to the Gal4 tag, assay the enzymatic functions of the ORF2p fragments independent of one another, and refine the sequence requirements for the minimally functional ORF2p molecule. The utility of this method is demonstrated through the identification and initial characterization of two novel sets of amino acids important to retrotransposition.

MATERIALS AND METHODS

Naming Conventions

All ORF2p fragments are named systematically. For C-terminally truncated fragments, the previously reported domain(s) is(are) used as the body of the name, followed by the number corresponding to the terminal amino acid as it would be in the full length ORF2p, with the deleted C-terminal ORF2p sequences denoted by a ∆. For example, the ORF2p fragment that contains the EN domain and Z domain that ends at ORF2p amino acid 490 is written as ENZ490∆. For N-terminally truncated ORF2p fragments, the core domains are again used as the name base. The name begins with a ∆ corresponding to the ORF2p sequence at the N-terminus missing from the construct, followed by the amino acid number representing the amino acid position marking the beginning of the fragment within the context of the full-length ORF2p. For example, a fragment that begins at amino acid 348 and contains the RT and Cys domains is written as ∆348RTCys. By convention, the Z domain is not included in the name of RT-containing fragments.

Cloning

Previously reported plasmids containing both codon optimized (L1PA1 Chang) [10] and wild-type (L1 lis) [11] ORF2 sequence were used as the templates for generation of truncated ORF2 PCR products and their subcloning into pcDNA 3.1/Hygro+ (Life Technologies) and pBind (Promega). ORF2 fragments containing inactivating D205A and H230A mutations within the EN sequence [1,12] were cloned using previously reported expression plasmid containing codon optimized ORF2 sequence [10] to generate ENn expression plasmids designed to produce protein fragments with non-functional endonucleases. To generate untagged ORF2p fragment expression constructs, an NheI restriction site-Kozac -ATG and TGA-HindIII restriction site were added 5′ and 3′ of the ORF2 DNA sequence of interest. PCR products were digested with NheI and HindIII and cloned into pcDNA3.1/Hygro+ (Life Technologies). To generate Gal4 tagged or VP16 tagged (neomycin resistant plasmid) ORF2p fragment expression constructs, PCR primers were designed using the Flexi Vector primer design tool (Promega). These PCR primers added a 5′ SgfI restriction site and a 3′ PmeI restriction site containing a valine (V) codon (GTT) and a stop codon (TAA) to the ORF2 sequence of interest. PCR products were digested with SgfI/PmeI blend (Promega: catalogue number R1852) and cloned into the pBind vector (Promega Checkmate Mammalian Two Hybrid System: catalogue

number C934A). 3xFLAG-tagged, Gal4-tagged ENZ and RTCys constructs were generated using overlapping PCR using the above-described codon optimized Gal4-tagged ORF2 plasmids (primers available upon request).

Site-directed mutagenesis

PCNA binding mutants. ENZ490∆ and ∆348RTCys constructs with previously reported PCNA binding mutation (YY 414–415 AA) [5] were generated using the QuikChange Site-Directed mutagenesis kit (StrataGene: catalogue number 240205-5) per the manufacturer's protocol using the following primers:

5′-CCGAGATCCAGACCACCATCCGGGAGGCCG CCAAGCACCTGTACGCCAACAAGCTGG-3′ and 5′ -CCAGCTTGTTGGCGTACAGGTGCTTGGCGGC CTCCCGGATGGTGGTCTGGATCTCGG-3′

FF PCNA mutants. ENZ490∆ and EN347∆ constructs with (FF 273–274 AA) in the putative PCNA binding domain were generated using the QuikChange Site-Directed mutagenesis kit (StrataGene) per the manufacturer's protocol using the following primers: 5′-CGAGATGAAGGCCGAGATCAAGATGGCCGC CGAGACCAACGAGAACAAGGACACC-3′ and 5′-GGTGTCCTTGTTCTCGTTGGTCTCGGCGGCCA TCTTGATCTCGGCCTTCATCTCG-3′

WD mutants. The same approach as above was used to generate L1[13], L1neo[13], ORF2, ENZ490∆, and ∆348RTCys constructs containing WD 288–289 AA mutations and ENZ490∆ constructs containing a W288A or D289A mutation using the following pairs of primers: 5'-GGACACCACCTACCAGAACCT GGCC GCCGCCTTCAAGGCCGTGTGCC-3' and 5'-GGCAC ACGGCCTTGAAGGCGGCGGCCAGGTTCTGGTA GGTGGTGTCC-3'; 5'-GGACACCACCTACCAGAAC CTGGCCGACGCCTTCAAGGCCGTGTGCC-3' and 5'-GGCACACGGCCTTGAAGGCGTCGGCCAGGTT CTGGTAGGTGGTGTCC-3'; 5'-GGACACCACCTAC CAGAACCTGTGGGCCGCCTTCAAGGCCGTGTG CC-3' and 5'-GGCACACGGCCTTGAAGGCGGCC CACAGGTTCTGGTAGGTGGTGTCC-3'.

Untagged BAR

Figure 2. HeLa cells were maintained in MEM supplemented with 1% sodium pyruvate, L-glutamate, NEAA and 10% FBS as previously described [14]. 1 million cells were seeded 16–18 h prior to transfection in T75 flasks. 0.1 µg ENZ490∆ and 0.7 µg ∆348RTCys plasmids were cotransfected with 1 µg previously reported Alu reporter construct [15] using 6 µl Plus reagent (Life Technologies) in 200 µl DMEM and 8 µl Lipofectamine reagent in 92 µl DMEM. Empty vector (pcDNA3.1/Hygro+) was used to ensure that the control reactions had equal amounts of DNA in all transfections. Cell culture media was supplemented with 0.45 mg/ml neomycin ∼24 h post-transfection. Colonies were stained after 2 weeks with Neomycin selection with crystal violet solution (0.2% crystal violet, 5% acetic acid,

2.5% isopropanol) and counted with Oxford Optronics Col-Count. Statistical significance assessed using Student's *t*-test for paired samples (*n* = 3), with error bars denoting standard deviation.

Gal4-tagged BAR and ORF2

Figures 3–9, 12, 13, 16, Supplementary Figure S6. Five hundred thousand HeLa cells were seeded 16–18 h prior to transfection in T75 flasks. 0.4 μg EN-containing plasmid and 0.4 μg Δ348RTCys plasmid were cotransfected with 0.4 μg previously reported Alu reporter construct (15) transfected using 6 μl Plus reagent (Life Technologies) in 200 μl DMEM and 8 μL Lipofectamine reagent in 92 μl DMEM. Empty vector (pBind) was used to ensure that control reactions had equal amounts of DNA in all transfections. Colonies were selected, stained, and counted as above. Statistical significance assessed as above.

Figure 10. One million HeLa cells were seeded 16–18 hours prior to transfection in T75 flasks. 0.1 μg ENn-containing plasmid or Cryptic-containing plasmid and 0.7 μg Δ348RTCys plasmid were cotransfected with 1 μg previously reported Alu reporter construct (15) transfected with 8 μl Lipofectamine reagent (Life Technologies) and 6 μl Plus reagent (Life Technologies). Empty vector used to ensure controls had equal amounts of DNA in all transfections. Colonies were selected, stained, and counted as above. Statistical significance assessed as above.

Toxicity assay

Supplementary Figures S4 and S10. ORF2p EN acute toxicity assay was performed in HeLa cells as previously described (8). Five hundred thousand cells were seeded 16–18 h prior to transfection in T75 flasks. 1 μg EN or ENn plasmid and 1 μg previously described pIRES2-GFP plasmid (source of neomycin resistance) cotransfected with 8 μl Lipofectamine reagent (Life Technologies) and 4 μl Plus reagent (Life Technologies). Empty vector used to ensure controls had equal amounts of DNA in all transfections. Colonies were selected, stained, and counted as above. Statistical significance assessed as above.

L1 retrotransposition

Figure 14. Modified from (16). HeLa cells maintained as previously described. 500 000 cells seeded 16–18 h prior to transfection in T75 flasks. 1 μg L1 construct was transfected in 4 μl Plus reagent (Life Technologies) and 8 μl Lipofectamine reagent (Life Technologies).

Immunoblot analysis

Transfection, cell culture and total protein harvest. Two million cells seeded 16–18 h prior to transfection in T75 flasks. 6 μg of appropriate expression construct or appropriate empty vector (control) were transfected with 24 μl Lipofectamine reagent (Life Technologies) and 12 μl Plus reagent (Life Technologies) (17). For coexpression experiments (Supplementary Figures S4 and S7), 4 μg of appropriate expression constructs were transfected separately and

together, with empty vector used as filler to bring total DNA transfected to 8 μg. Approximately 24 h post-transfection, cells were washed 1× with phosphate buffered saline (PBS) and then harvested in 500 μl Total lysis buffer (TLB: 50 mM Tris, 150 mM NaCl, 10 mM EDTA, 0.5% SDS, 0.5% Triton-X, pH 7.2) supplemented with 10 μl/ml of Halt protease inhibitor cocktail, phosphate inhibitor cocktail 2, and phosphate inhibitor cocktail 3 (Sigma). After one round of freeze (−80°C)/thaw on ice, cells were sonicated with a Microson XL-2000 sonicator (Misonix) 3× (10 s sonication/10 s rest on ice), and cell lysates were then centrifuged at 4°C at 14 000 rpm for 15 min. Protein concentrations of cleared cell lysates were determined using BioRad protein assay (Bradford method).

Western blot analysis.

Figures 2 and 3. 30 μg of total cell lysate were boiled in equal volume of Laemmli buffer with 3.4% β-mercaptoethanol. Samples were fractionated on 3–8% Tris-acetate gels (Life Technologies) and transferred to nitrocellulose membranes using the iBlot system (Life Technologies). Membranes were blocked in PBS-Tween (PBS, 0.1% Tween) with 5% blotting-grade blocker (Bio-Rad) and incubated with primary antibodies overnight at 4°C. Untagged ENZ490Δ protein detected with previously reported custom polyclonal ORF2p antibodies (8) (1:500 in 3% blotting grade blocker/PBS-Tween). Untagged Δ348RTCys protein was detected with previously reported custom polyclonal ORF2p antibodies (1:500 in 3% blotting grade blocker/PBS-Tween) (8). Membranes were washed following overnight primary antibody incubation 3× in PBS-Tween for 5 min. Secondary antibody (Santa Cruz) applied for 1 h at 25°C (1:5000 in 3% blotting grade blocker/PBS-Tween). As appropriate, either HRP-donkey anti-goat or HRP-goat anti-rabbit were used. Western blots were developed using the Immun-Star WesternC kit (BioRad). Images were captured using a BioRad Gel Doc XR+ imager. GAPDH (Santa Cruz: sc-25778) was used as loading control (1:3000 in 3% blotting grade blocker/PBS-Tween). Statistical significance of ORF2p or protein fragment band signal intensity differences relative to GAPDH loading control assessed using Student's *t*-test for paired samples (*n* = 3), with error bars representing standard deviation.

Figures 4–8, 10, 12, 13, 16, Supplementary Figures S5–S8. 30 μg total cell lysate were heated at 85°C for 5 min in Laemmli buffer without β-mercaptoethanol supplementation (17). Samples were fractionated, transferred, and blocked as described above, and incubated with primary antibodies overnight at 4°C. All Gal4 tagged constructs detected using anti Gal4 DNA binding domain antibodies (Santa Cruz: sc-577: 1:1000 dilution). Membranes were washed following overnight primary antibody incubation 3× in PBS-Tween for 5 min. Secondary antibody applied for 1 h at 25°C (HRP-goat anti-rabbit 1:5000 in 3% blotting grade blocker/PBS-Tween) western blots were developed, documented, and analysed as above.

Amino acid sequence alignment

Amino acid sequence alignment performed using MegAlign software (DNASTAR v. 10.0.1). Sequences aligned using clustal W method relative to the human ORF2p sequence. Amino acid sequence group match strength indicated by above histogram (>50% sequences match). Sources of ORF2p and ORF2p-like molecule sequence with amino acid numbers used for alignment and GenBank accession numbers are as follows: *Mus musculus* (AAA39398): amino acids 262–536 *Rattus norvegicus* (AAB41224): amino acids 265–363 *Caenorhabditis elegans* (AAC72298): amino acids 357–604 *Pan troglodytes* (ABE73458): amino acids 237–510 *Sus scrofa* (ABR01162): amino acids 237–509 *Canis lupus* (BAA25253): amino acids 238–510 *Chlorella vulgaris* (BAA25763): amino acids 257–548 *Danio rerio* (BAC82613): amino acids 252–520 *Ciona intestinalis* (BAC82626): amino acids 241–515 *Anopheles gambia* (BAC82628): amino acids 234–500 *Takifugu rubripes* (BAD04856): amino acids 234–499 *Tetraodon nigroviridis* (BAD04858): amino acids 234–499 *Bos grunniens* (ELR58410): amino acids 237–509

L1 element amplification protocol (LEAP)

Adapted from (18,19). Briefly, HeLa cells were maintained as described above. Cells were seeded and transfected as described above with 6 μg of appropriate gal4-tagged ORF2 expression plasmids. Cells were washed as described above and harvested in 5 ml PBS. Cells were pelleted at 6000 rpm for 5 min at 4°C and supernatant removed. Cells were lysed in 500 μl of LEAP lysis buffer (1.5 mM KCl, 2.5 mM $MgCl_2$, 5 mM Tris–Cl, 1%DOC, 1% Triton X-100, EDTA and RNAsin), incubated on ice for 5 min, and briefly centrifuged to remove debris. Supernatant was layered on top of an 8.5% sucrose cushion that was layered on top of a 17% sucrose cushion. Samples were centrifuged at 36 500 rpm for 2 h at 4°C. Supernatant was removed and pellet resuspended in 100 μl deionized water supplemented with HALT protease inhibitor and RNAsin. Protein concentration was assessed using Bradford assay and samples brought to equal protein concentrations using glycerol. RTPCR reaction was performed using 900 ng of sample as described in (18,19). For detection of Actin and ORF2 RNA, Reverse Transcriptase System (Promega: A3500) was used according to manufacturer's instructions (Actin primers available upon request). Western blot analysis on LEAP samples performed as described above.

RESULTS

ORF2p fragments can reconstitute the function of full length ORF2p as measured by successful Alu retrotransposition

We hypothesized that the intervening sequence between the EN and Z domains has some functional importance to retrotransposition. We reasoned that the importance of said ORF2p sequence could be interrogated systematically by separating the ORF2p molecule into two fragments and by supplying these fragments with one another to drive Alu retrotransposition. We call this approach Bipartile Alu Retrotransposition (BAR).

To test the functionality of this approach, we generated two expression plasmids (ENZ490Δ and Δ348RTCys) containing truncated fragments of codon optimized ORF2 DNA sequence (Figure 2A). The ENZ490Δ plasmid was designed to contain the DNA sequence encoding for the ORF2p EN and Z domains, beginning with the N-terminus of the ORF2p and ending at ORF2p amino acid 490. The Δ348RTCys plasmid was designed to express an ORF2p fragment that began at ORF2p amino acid 348 and contained the ORF2p Z, RT and Cys domains, ending at the C-terminus of the ORF2p molecule. All subsequently generated constructs are named in a similar manner to the above constructs: the core name of each construct is derived from the previously described domains located in the ORF2p fragment the construct is designed to express. By convention, the Z domain is not included in the name of RT-containing constructs. The N- and C-terminal truncation positions are designated with a Δ sign included before or following the core name of the construct, respectively. The amino acid number corresponds to the beginning or the end of the fragment at the site of truncation relative to the full length ORF2p molecule. These break points were chosen to (i) physically separate the EN and RT domains, (ii) have overlapping sequence (the Z domain) when the two ORF2p fragments are expressed, and (iii) have only one fragment containing the protein sequence between the EN and Z domains (EN490Δ).

Expression of the above described ORF2p fragments in HeLa cells was assessed 24 h after the cells were transiently transfected with the ENZ490Δ and Δ348RTCys expression constructs. Western blot analysis using custom polyclonal ORF2p antibodies (8) detected both truncated ORF2p fragments (Figure 2B). Following confirmation of ENZ490Δ and Δ348RTCys protein expression, their ability to drive Alu retrotransposition was tested using transient cotransfection of HeLa cells with the ENZ490Δ, Δ348RTCys, and the previously described neomycin resistance gene (Neo[r])-tagged Alu expression constructs (Figure 2C). Low levels (∼50 colonies) of Alu retrotransposition were observed only when both the ENZ490Δ and the Δ348RTCys expression plasmids were transfected in HeLa cells (Figure 2D). No retrotransposition events were observed when the Alu reporter construct was cotransfected into HeLa cells with either the ENZ490Δ or the Δ348RTCys constructs alone (Figure 2D). This approach demonstrated that, albeit inefficiently, the ORF2p fragments, each lacking essential portions of the ORF2p molecule, can drive Alu retrotransposition when expressed from different RNAs. This proof of principle result allowed us to proceed with the manipulation of the ORF2p fragments independent of one another and to test the effects of these manipulations on BAR.

An N-terminal Gal4 tag significantly improves BAR efficiency

To improve the efficiency of BAR, we introduced a Gal4 DNA binding domain at the N-terminus of each of the two above-described ORF2p fragments (Figure 3A). The Gal4 tag was chosen so that all protein fragments could be detected with a single antibody. Also, it has been sug-
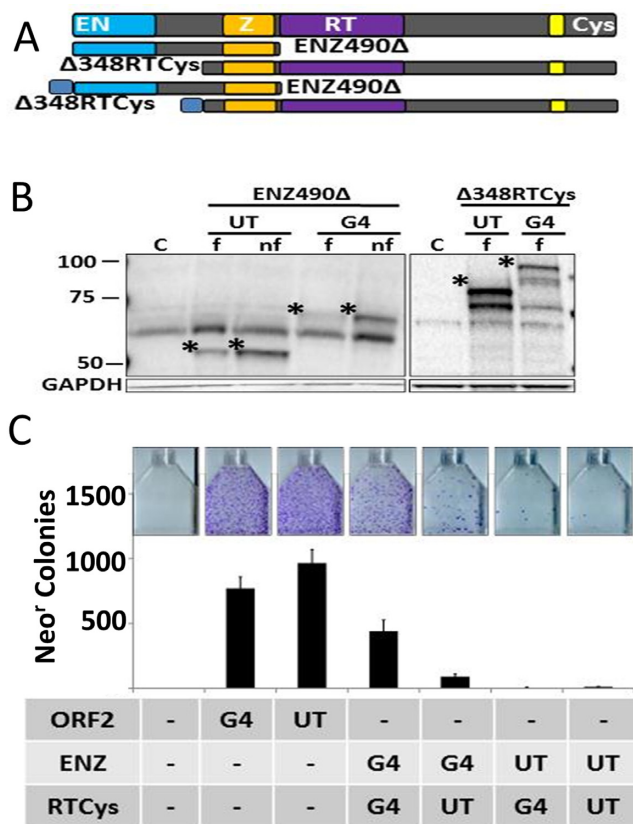
Figure 3. Addition of an N-terminal Gal4 DNA binding domain to protein fragments significantly improves BAR efficiency. (**A**) Schematic representation of the ORF2 fragments N-terminally fused with a Gal4 DNA binding domain (gray–blue square). ORF2 fragments generated using codon optimized ORF2 DNA sequence. (**B**) Western blot analysis of protein expression in HeLa cells with custom polyclonal ORF2p antibodies (Goat 123 for ENZ490Δ, Rabbit 960 for Δ348RTCys). Bands of expected sizes for untagged (UT) and tagged (G4) ENZ490Δ and ENZ490Δn (catalytically inactive due to D205A and H230A mutations) as well as untagged Δ348RTCys and tagged Δ348RTCys are denoted by black asterisks. Functional EN-fragment denoted by (f), while nonfunctional is denoted by (nf). Control lane (C) is total cell lysate from HeLa cells transfected with empty vector. Molecular weight markers are indicated on the left. GAPDH used as loading control. Bands of smaller molecular weight may represent truncated/processed/degraded ORF2p fragment products. (**C**) Addition of Gal4 tag significantly improves BAR efficiency. Tagged and untagged ORF2p and its fragments are indicated as G4 or UT. Error bars denote standard deviation (*n* = 3). Representative flasks are shown above corresponding graph bars.

gested that the ORF2p functions as a dimer (20,21). Gal4 is known to function as a dimer (22): therefore we reasoned that the addition of this tag to the N-termini of ORF2p fragments may facilitate interactions that usually take place within the full-length molecule. Both tagged and untagged ENZ490Δ and Δ348RTCys protein fragments were detected *via* western blotting with custom polyclonal ORF2p antibodies following transient transfection of HeLa cells with their respective expression plasmids (Figure 3B). This approach demonstrated that the addition of the Gal4 tag did not significantly alter expression of either ENZ490Δ or Δ348RTCys proteins.

To characterize the effect of this tag on the ability of these two ORF2p fragments to support Alu retrotranspo-

sition, all possible combinations of constructs expressing Gal4 -tagged and untagged ENZ490Δ and Δ348RTCys proteins were tested in BAR. A combination of Gal4 tagged versions of both ENZ490Δ and Δ348RTCys fragments gave the most robust BAR signal, yielding retrotransposition levels at 50–60% of Alu retrotransposition driven by the full-length Gal4-tagged ORF2p (Figure 3C: G4 ORF2 versus G4 ENZ and G4 RTCys). Further, cotransfection of the Gal4-tagged ENZ490Δ and Δ348RTCys constructs resulted in a 60-fold increase in the BAR signal (colony number) relative to the BAR levels supported by the proteins produced by the untagged versions of these constructs (Figure 3C). Although we were unable to determine if the two Gal4-tagged protein fragments directly interact using co-immunoprecipitation of the Gal4-tagged ENZ-containing proteins with an N-terminally 3xFLAG-tagged Gal4-tagged RTCys-containing protein (Supplementary Figure S1), the addition of this tag significantly improved BAR efficiency (Figure 3C). Interestingly, tagging just the ENZ490Δ fragment with the Gal4 DNA binding domain improved the BAR signal by 12-fold over baseline BAR levels yielded through the use of untagged ENZ490Δ and Δ348RTCys expression constructs (Figure 3C: G4 ENZ490Δ and UT Δ348RTCys *vs.* UT ENZ490Δ and UT Δ348RTCys). The combination of the Gal4-tagged ENZ490Δ expression plasmid with the untagged Δ348RTCys expression plasmid produced retrotransposition levels reaching 10–15% of Alu retrotransposition driven by the Gal4-tagged full-length ORF2p (Figure 3C). In contrast, tagging only the Δ348RTCys fragment with the Gal4 DNA binding domain did not significantly improve BAR signal over baseline levels (Figure 3C: untagged fragments only). These results demonstrate that the N-terminal Gal4 tag significantly improved the BAR signal, allowing us to proceed with further characterization of sequence requirements within each ORF2p fragment for Alu retrotransposition.

The RTCys protein fragment may be the rate-limiting protein fragment in BAR

The above mentioned experiments were performed using plasmids containing codon optimized L1 ORF2 DNA sequences. To determine the influence of protein expression levels on BAR, we generated Gal4 tagged ENZ490Δ and Δ348RTCys expression plasmids containing wild-type ORF2 DNA sequence (Figure 4A). Protein fragments were expressed from both wild-type and codon optimized ORF2 DNA sequence-containing ENZ490Δ and Δ348RTCys expression plasmids. The protein fragments were detected following transient transfection of HeLa cells with corresponding expression constructs *via* western blot analysis with commercially available anti-Gal4 DNA binding domain antibodies (Santa Cruz) (Figure 4B). Following transient transfection of HeLa cells, wild-type ORF2 DNA sequence-containing ENZ490Δ and Δ348RTCys constructs expressed their respective proteins at ∼30% the levels of the proteins expressed from corresponding constructs made using codon optimized ORF2 DNA sequence (Figure 4C).
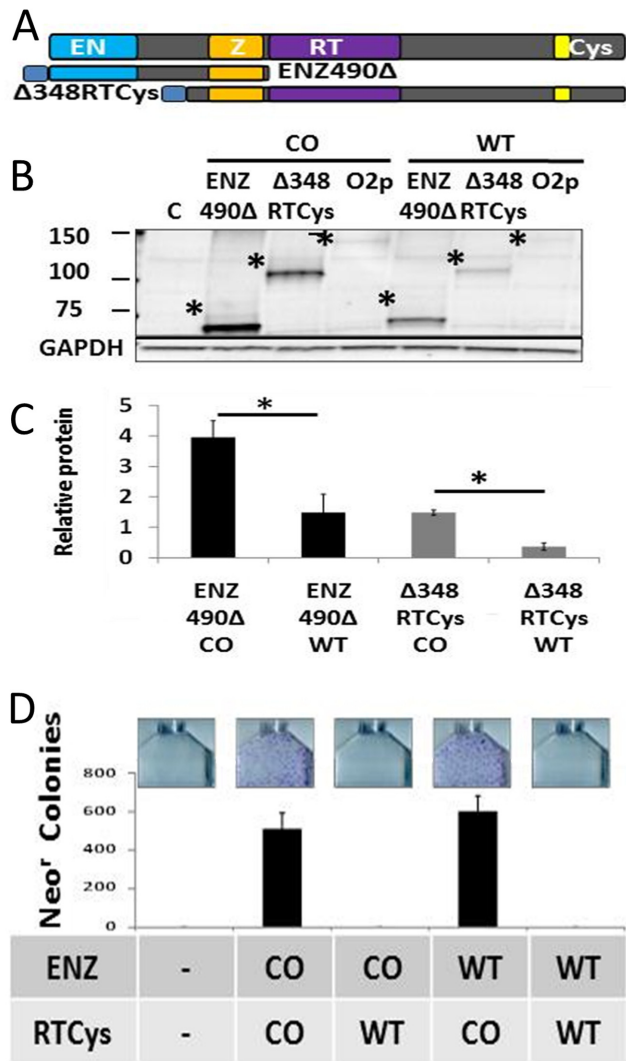
Figure 4. The RTCys fragment may be the rate-limiting protein fragment in BAR. (**A**) Schematic representation of the Gal4-tagged (grey-blue square) ORF2 fragments generated using both codon optimized and wild-type ORF2 DNA sequence. (**B**) Western blot analysis of ORF2p and its fragments originating from both codon optimized (co) and wild-type (wt) ORF2 DNA sequence using Gal4 antibodies (Santa Cruz). Bands of expected sizes are denoted by black asterisks. Control lane (C) is total cell lysate from HeLa cells transfected with empty vector. Molecular weight markers are indicated on the left. GAPDH used as loading control. (**C**) Quantitation of protein expression from ENZ490Δ and Δ348RTCys expression plasmids containing codon optimized or wilt-type ORF2 DNA sequence. Protein fragment expression normalized to GAPDH loading control. Error bars denote standard deviation ($n = 3$). Statistical significance assessed using Student's *t*-test for paired samples, with denoted by *. (**D**) BAR output using combination of ENZ and RTCys fragments shown in (A) generated using both codon optimized (CO) and wild-type (WT) ORF2 DNA sequence. Error bars denote standard deviation ($n = 3$). Representative flasks are shown above corresponding graph bars.

To test the effect of protein levels on Alu retrotransposition, the Gal4-tagged ENZ490Δ and Δ348RTCys expression plasmids containing wild-type or codon optimized ORF2 DNA sequence were tested in various combinations in BAR. The Δ348RTCys protein fragment expressed from the wild-type ORF2 DNA sequence did not support BAR
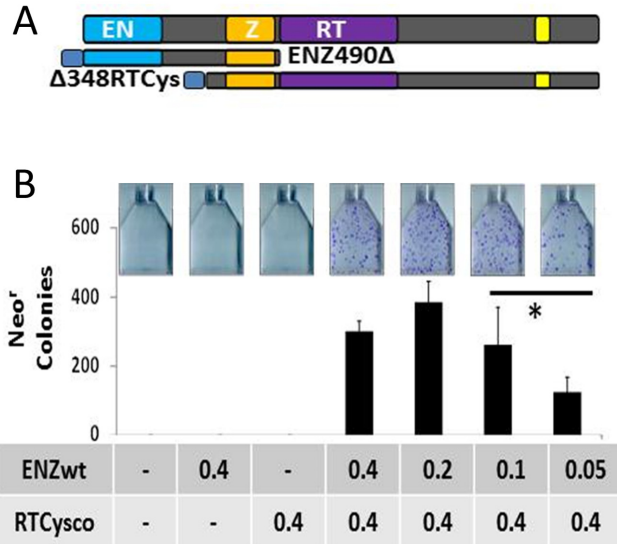

Figure 5. Dose effect of ENZ fragment on BAR efficiency. (**A**) Schematic representation of the Gal4-tagged (gray–blue square) ORF2 fragments. ENZ490Δ fragment was generated from wild-type ORF2 sequence, while Δ348RTCys fragment was generated using codon optimized ORF2 sequence. (**B**) BAR output using different amounts (0.4–0.05 μg) of the wild-type ENZ expression plasmid with 0.4 μg of the RTCys expression plasmid. Empty vector used as filler to ensure equal DNA amounts in transfections. Error bars denote standard deviation ($n = 3$). Statistical significance assessed using Student's *t*-test for paired samples with denoted by *. Representative flasks are shown above corresponding graph bars.

when co-expressed with ENZ490Δ fragments, irrespective of whether ENZ490Δ protein fragment was from wild-type or codon optimized ORF2 DNA sequence (Figure 4D). Despite the 3-fold difference in protein expression between the ENZ490Δ protein fragments derived from codon optimized *vs.* wild-type ORF2 DNA sequences, both supported BAR in HeLa cells at similar levels when coexpressed with the Δ348RTCys protein produced by the expression plasmid containing codon optimized DNA sequence (Figure 4D). Indeed, a significant reduction in BAR efficiency was observed only when the amount of ENZ490Δ expression plasmid containing wild-type DNA sequence was reduced to ∼13% of the Δ348RTCys expression plasmid containing codon optimized sequence (Figure 5B: 0.4 versus 0.05 μg). These data demonstrate that the Δ348RTCys ORF2p fragment may be the rate-limiting protein fragment in BAR.

The PCNA binding site in the ORF2p Z domain can be supplied in a non-contiguous fashion relative to the RT-containing ORF2p fragment in BAR

We next characterized the effect of a mutation known to disrupt the function of the full-length ORF2p on BAR. PCNA binding to the ORF2p was reported to be essential to L1 retrotransposition (5). The PCNA binding domain of the ORF2p is located in the Z domain, which is present in both the ENZ490Δ and Δ348RTCys protein fragments supporting BAR (Figure 6A). To test the effect of PCNA mutations on BAR, we generated ENZ490Δ and Δ348RTCys constructs containing PCNA Interacting Protein (PIP) mutations (YY 414–415 AA) within the Z do-
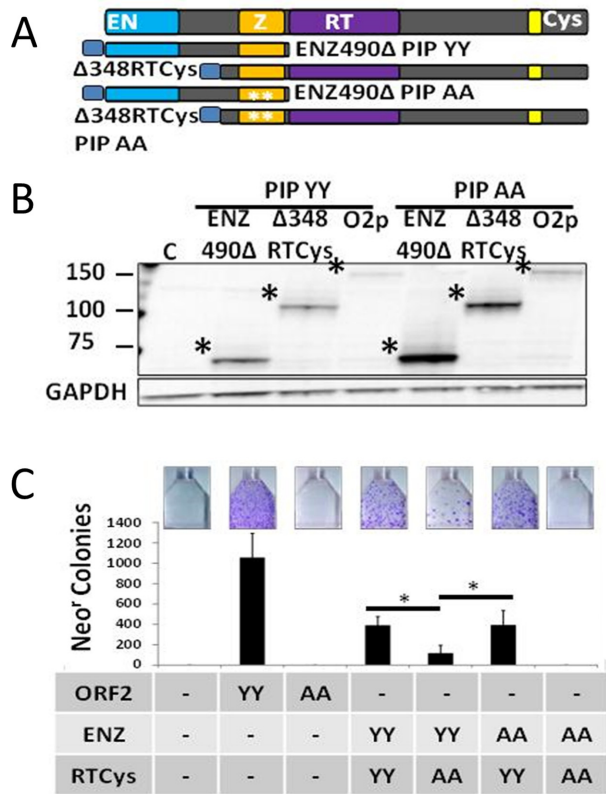
Figure 6. PCNA binding domain can be supplied *in trans* of the RTCys fragment in BAR. (**A**) Schematic representation of the Gal4-tagged (grey-blue square) ORF2 fragments generated with codon optimized ORF2 DNA sequence. Constructs were generated with nonfunctional PCNA binding domain (amino acids 414–415) and contain white asterisks at the approximate position of the mutations. (**B**) Western blot analysis of ORF2p and its fragments using Gal4 antibodies (Santa Cruz). Bands of expected sizes are denoted by black asterisks. Control lane (C) is total cell lysate from HeLa cells transfected with empty vector. Molecular weight markers are indicated on the left. GAPDH used as loading control. (**C**) BAR output in HeLa cells. Status of PCNA PIP indicated below corresponding graph bar for ORF2, ENZ, and RTCys (YY: functional, AA: mutant). Error bars denote standard deviation ($n = 3$). Statistical significance assessed using Student's *t*-test for paired samples with indicated by *. Representative flasks are shown above corresponding graph bars.

main (Figure 6A). An expression construct containing the full-length ORF2 sequence with the same mutation was also generated to serve as a control. Functional and PIP mutant ENZ490Δ protein fragments, Δ348RTCys protein fragments, and full-length ORF2p were expressed in HeLa cells transiently transfected with their respective expression constructs as detected by western blot analysis using anti-Gal4 antibodies (Santa Cruz) (Figure 6B). Consistent with the reported effect of the PIP mutations on L1 retrotransposition (5), mutation of the YY PCNA binding site in the full-length ORF2p abolished its ability to drive Alu retrotransposition in HeLa cells (Figure 6C). Similarly, ENZ490Δ and Δ348RTCys protein fragments containing PIP mutations did not support BAR when coexpressed in HeLa cells (Figure 6C). In contrast, functional and YY mutant ENZ490Δ protein fragments supported BAR at similar efficiencies when coexpressed with PCNA binding competent

Δ348RTCys protein fragment (Figure 6C: ENZ YY and RTCys YY *vs*. ENZ AA and RTCys YY). Importantly, a significant (25% of BAR supported by ENZ YY and RTCys YY) BAR signal was observed when the functional ENZ490Δ protein fragment was coexpressed with the PIP mutant Δ348RTCys protein fragment in HeLa cells (Figure 6C). These data support that PCNA binding can be supplied in a non-contiguous fashion relative to the RT-containing ORF2p fragment in BAR by the ENZ490Δ protein fragment.

Predicted FF PCNA binding domain within the ORF2p sequence between the EN and Z domains is required for Alu retrotransposition

To validate BAR as a tool for identification of novel L1 ORF2p sequences important for retrotransposition, we chose a an area of the ORF2p molecule outside of the previously described ORF2p domains that had strong sequence identity to known PCNA interaction motifs (FF PCNA), with essential core FF at ORF2p amino acids 273–274 (Supplementary Figure S2). We hypothesised that this putative FF PIP domain in the ORF2p region between the EN and Z domains may also be important for retrotransposition.

Using site directed mutagenesis, we engineered a codon optimized construct designed to express Gal4-tagged ENZ490Δ protein fragment that has the essential phenylalanine pair in the putative PCNA mutated to alanines (ENZ490Δ FF:AA Mutant) (Figure 7A). The same FF to AA mutations were also introduced into the full-length contiguous ORF2p. Western blot using anti-Gal4 antibodies (Santa Cruz) determined that ORF2p and ENZ490Δ protein containing mutated putative FF PCNA motif were expressed at similar levels to their functional equivalents following transient transfection of their expression constructs in HeLa cells (Figure 7B). The FF 273–274 AA mutation in the predicted PCNA binding motif abolishes BAR (Figure 7C: ENZ490Δ AA and Δ348RTCys) as well as retrotransposition driven by the full-length ORF2p (Supplementary Figure S3). The same result was observed with the EN347Δ construct (please see below). These data demonstrated that as with the previously reported PIP mutation (Figure 6), BAR accurately predicted the retrotransposition potential of the full-length linear ORF2p, and can identify novel amino acids important to retrotransposition.

Overlap in the Z domain between the ENZ490Δ and Δ348RTCys ORF2p fragments is not necessary for BAR

We next tested whether BAR is useful for the investigation of the importance of the ORF2p sequence between the EN and Z domains to retrotransposition. To determine whether the overlap in the Z domain introduced by the ENZ490Δ and Δ348RTCys protein fragments was necessary for BAR, the EN347Δ construct was designed to reconstitute the linear ORF2p sequence when combined with the Δ348RTCys construct without any gaps or overlaps. Additionally, the EN264Δ and EN289Δ expression constructs were designed to introduce an 83 or 58 amino acid gap in the ORF2p sequence when cotransfected with the
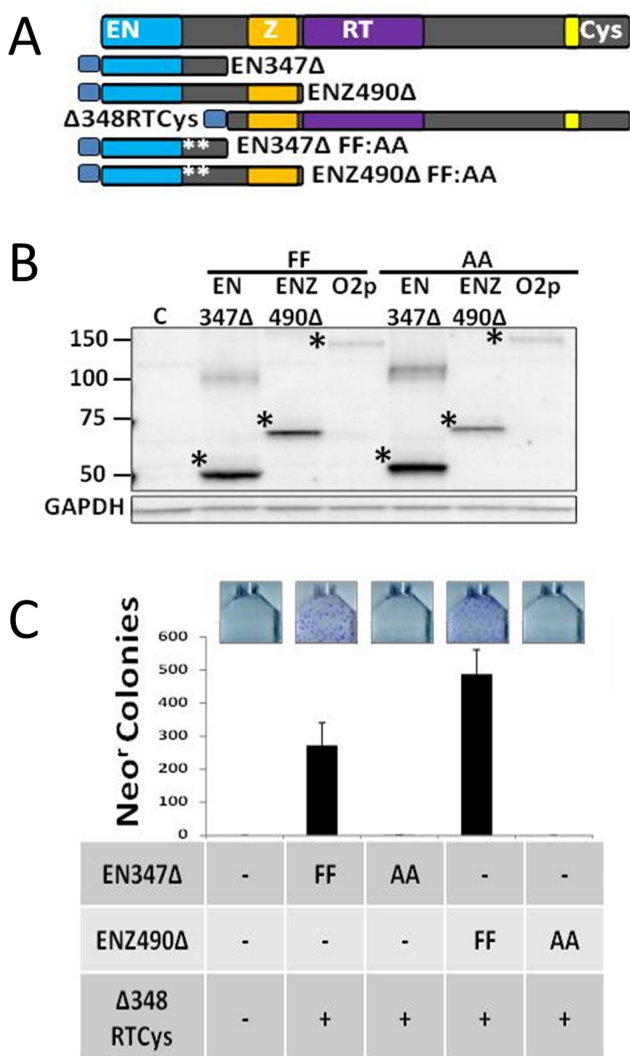
Figure 7. Mutation of predicted FF PCNA binding domain (FF PCNA) abolishes BAR. (**A**) Schematic representation of the Gal4-tagged (grey-blue square) ORF2 fragments generated using codon optimized ORF2 DNA sequence. EN-containing constructs were generated with nonfunctional putative FF PCNA (amino acids 273–274) with the approximate location of the mutations denoted by white asterisks. (**B**) Western blot analysis of ORF2p and its fragments using Gal4 antibodies (Santa Cruz). Bands of expected sizes are denoted by black asterisks. Control lane (C) is total cell lysate from HeLa cells transfected with empty vector. Molecular weight markers indicated at left. GAPDH used as loading control. (**C**) BAR output in HeLa cells. Status of EN-containing fragment putative FF PCNA denoted below corresponding graph column. Error bars denote standard deviation (*n* = 3). Representative flasks are shown above graph bars.

Figure 8. Overlap in the Z domain between ORF2 fragments is not necessary for BAR. (**A**) Schematic representation of the Gal4-tagged (grey-blue square) ORF2 fragments generated from codon optimized ORF2 DNA sequence. The EN264Δ fragment contains the EN domain and 25 amino acids between the EN and Z domains (amino acids 1 through 264). The EN289Δ fragment contains the EN domain and 50 amino acids between the EN and Z domains (amino acids 1 through 289). The EN347Δ fragment contains the EN domain and the sequence between the EN and Z domains (amino acids 1 –347). (**B**) Western blot analysis of ORF2p and its fragments using Gal4 antibodies (Santa Cruz). Bands of expected sizes are denoted by black asterisks. Control lane (C) is total cell lysate from HeLa cells transfected with empty vector. Molecular weight markers are indicated on the left. GAPDH used as loading control. (**C**) BAR output in HeLa cells. Error bars denote standard deviation (*n* = 3). Representative flasks are shown above corresponding graph bars.

Δ348RTCys construct. All three constructs were generated using codon optimized ORF2 DNA sequence. All truncated ORF2p fragments were expressed in HeLa cells following transient transfection with respective expression constructs as detected *via* western blot analysis with anti-Gal4 antibodies (Santa Cruz) (Figure 8B).

We have previously reported that the ENZ490Δ fragment does not cause toxicity in mammalian cells compared to the ORF2p fragment containing only the EN domain (EN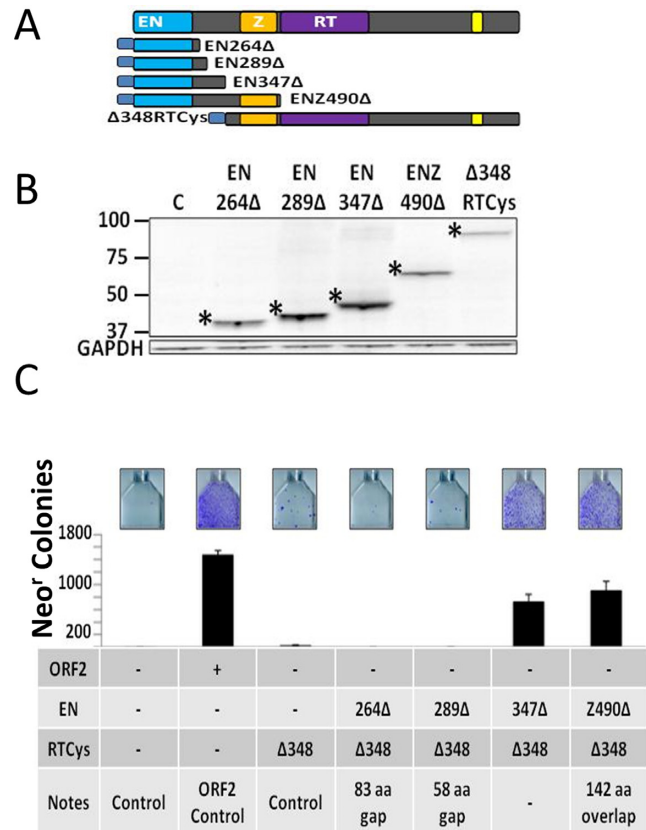239Δ) (8). This observation suggests that the EN-containing ORF2p fragments including varying amounts of the ORF2p sequence C-terminal to the EN domain may exhibit varying amounts of cytotoxicity, which could be important for interpreting BAR results. Therefore, these EN-containing protein fragments were first assayed for their ability to cause cytotoxicity when expressed in HeLa cells as previously described (8). In addition to plasmids designed to express catalytically active EN-containing protein fragment, expression plasmids engineered to express catalytically inert EN-containing protein fragments were generated to serve as controls. This was achieved by mutating the previously described essential aspartate (D205) and histidine (H230) residues into alanine residues (1,23). Transient transfection of HeLa cells with the functional EN264Δ expression plasmid resulted in cytotoxicity similar to that observed with the EN239Δ expression plasmid (Supplementary Figure S4: functional *vs.* non-functional).
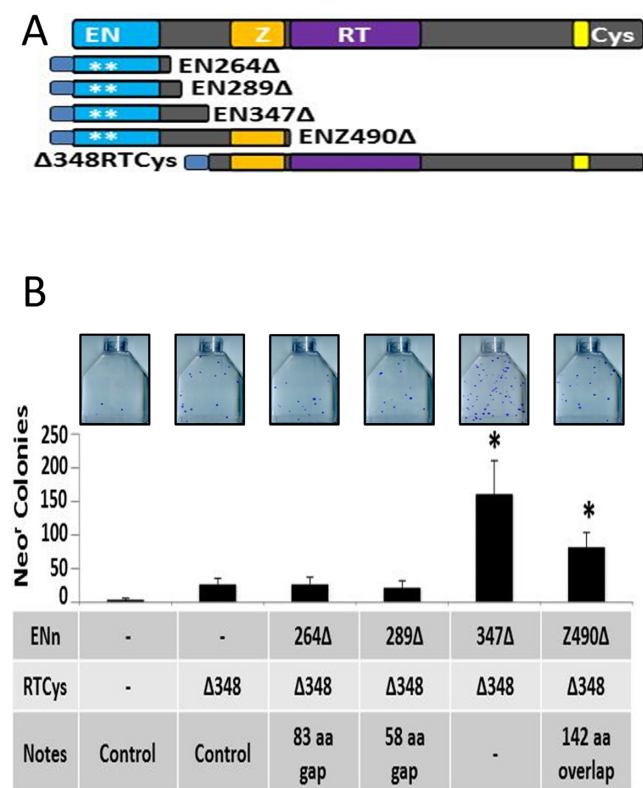
Figure 9. Catalytically inactive EN-containing fragments can increase residual Alu retrotransposition driven by the RTCys-containing fragment. (**A**) Schematic representation of the Gal4-tagged (grey-blue square) ORF2 fragments generated using codon optimized ORF2 sequence. (**B**) BAR output in HeLa cells. Error bars denote standard deviation (*n* = 3). Statistical significance assessed using Student's *t*-test for paired samples (* versus Δ348RTCys control). Representative flasks are shown above corresponding graph bars.

No cytotoxicity using the same experimental conditions was associated with the expression of EN289Δ, EN347Δ, or ENZ490Δ protein fragments in HeLa cells (Supplementary Figure S4). Using these EN-containing constructs, the BAR assay demonstrated that the overlap in the Z domain is not necessary for BAR (Figure 8C: EN347Δ and Δ348RTCys). However, the 83 amino acid gap (Figure 8C: EN264Δ and Δ348RTCys) and 58 amino acid gap (Figure 8C: EN289Δ and Δ348RTCys) are not compatible with BAR. The RTCys fragment was also cotransfected with EN-containing fragments that do not (EN289Δ) and do (ENZ490Δ) support BAR for Western Blot analysis of coexpression and steady-state protein levels (Supplementary Figure S5). Steady state levels of the RTCys-containing protein fragment did not appear to increase upon cotransfection. These results support that the ORF2p sequence between the EN and Z domains plays an important role in retrotransposition.

Catalytically inactive EN-containing fragments increase Alu retrotransposition driven by the RT-containing ORF2p fragment

To test the hypothesis that the ORF2p sequence between the EN and Z domains is important for retrotransposition, we

assessed whether the non-functional EN-containing fragments shown in Supplementary Figure S4 have any effect on the residual EN-independent retrotransposition of Alu driven by the Δ348RTCys protein fragment (Figure 8C: RT-Cys Control). Previously reported PCR analysis of genomic DNA extracted from cells harbouring these retrotransposition events demonstrated that these events resulted in spliced neo RNA being converted into cDNA that was then integrated into the host cell genome, consistent with *bona fide* retrotransposition events (Supplementary Figure S12) (11). These non-functional constructs express catalytically inactive EN-containing protein fragments that are unable to generate the 3′ OH necessary for EN-dependent TPRT (Figure 9A). Thus, this approach allows for testing the importance of the sequence between the Z and EN domains to retrotransposition independent of the EN catalytic activity. This approach demonstrated that two non-functional EN-containing ORF2p fragments, EN347Δ and ENZ490Δ, significantly increased residual Alu retrotransposition driven by the Δ348RTCys protein fragment; activity that is possibly reliant upon the suggested weak endonuclease activity observed in EN-mutated ORF2p (1). A 6-fold and 3-fold increase in Alu retrotransposition driven by the Δ348RTCys protein fragment was detected when Δ348RTCys was coexpressed with the non-functional EN347Δ or ENZ490Δ protein fragments, respectively relative to Alu retrotransposition driven by Δ348RTCys alone (Figure 9B). The other two EN-containing protein fragments (EN264Δ and EN289Δ) had no detectable effect on Alu mobilization driven by the Δ348RTCys protein fragment (Figure 9B). These data further support that the ORF2p sequence between the EN and Z domains may serve an important function in retrotransposition.

The ORF2p sequence between the EN and Z domains contains evolutionarily conserved amino acids and is sufficient to increase Alu retrotransposition driven by the RTCys fragment

We hypothesized that the ORF2p sequence between the EN and Z domains, which we termed Cryptic, can influence retrotransposition independent of the structural context of the EN domain. To test this hypothesis, we generated a Cryptic347Δ construct designed to yield a Gal4-tagged fragment only containing the Cryptic sequence (amino acids 240–347 of the full length ORF2p) (Figure 10A). We also generated a CrypticZ490Δ construct engineered to express a Gal4-tagged ORF2p fragment containing the Cryptic sequence and the Z domain (amino acids 240 through 490 of the full-length ORF2p) (Figure 10A). These expression plasmids were generated using codon optimized ORF2 DNA sequence.

Western blot analysis with anti-Gal4 antibodies (Santa Cruz) determined that the above-described ORF2p fragments were expressed in HeLa cells transiently transfected with their respective expression plasmids (Figure 10B). A significant increase in residual Alu retrotransposition driven by the Δ348RTCys fragment was observed with the addition of either Cryptic347Δ or CrypticZ490Δ fragments (Figure 10C). This increase was similar to the one observed when the Δ348RTCys expression plasmid was cotransfected with the catalytically inert EN347Δ and ENZ490Δ
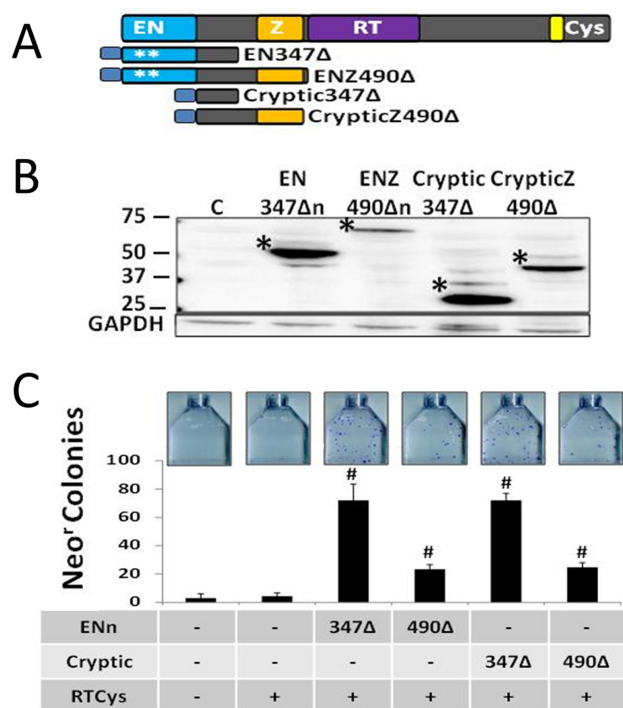
Figure 10. The Cryptic Sequence increases residual Alu retrotransposition supported by the RTCys Fragment. (**A**) Schematic representation of the Gal4-tagged (gray–blue square) ORF2 fragments generated using codon optimized ORF2 sequence. The Cryptic347Δ fragment contains the Cryptic sequence alone (amino acids 240 through 347). The CrypticZ490Δ fragment contains the Cryptic sequence and Z domain (amino acids 240–490). (**B**) Western blot analysis of ORF2p fragments using Gal4 DNA binding domain antibodies (Santa Cruz). Bands of expected sizes are denoted by black asterisks. Control lane (C) is total cell lysate from HeLa cells transfected with empty vector. Molecular weight markers indicated at left. GAPDH used as loading control. (**C**) BAR output in HeLa cells. Error bars denote standard deviation ($n = 3$). Statistical significance assessed using Student's t-test for paired samples (* versus Δ348RTCys control). Representative flasks are shown above corresponding graph bars.

expression plasmids, respectively (Figure 10C). These data demonstrate that the Cryptic protein sequence alone is important for BAR, and thus may be important for Alu mobilization driven by the full-length ORF2p as well as for L1 retrotransposition.

We hypothesized that any amino acids in the Cryptic sequence essential to BAR and retrotransposition in general are likely to be conserved in ORF2p and ORF2p-equivalent molecules in an array of taxonomically diverse species. The amino acid sequences from ORF2p or ORF2p-like proteins from a variety of species were aligned using the MegAlign software (clustal W method). Instead of using the entire ORF2p or ORF2p-like protein sequence, we utilized only the sequence immediately C-terminal to the EN domain through the sequence immediately N-terminal to the RT domain (Human L1 ORF2p amino acids 240 through 510), as annotated in the downloaded protein sequence files (UniProt). This alignment revealed an area of strong conservation that includes a tryptophan and aspartic acid (WD) pair, corresponding to human L1 ORF2p amino acids 288 and 289 (Figure 11). These amino acids have never been examined for their importance to retrotransposition,

and are located outside of the ORF2p EN domain with known crystal structure (23). They are also outside of the core RT domain, for which there is a predicted structure that is shared by most RTs (2,24–26). Due to the previously reported role of tryptophan in protein–protein interactions (27) and the evolutionary conservation of these amino acids, we hypothesized that these amino acids are important for ORF2p-dependent retrotransposition.

Highly conserved WD amino acids in the Cryptic sequence of the ORF2p are required for BAR, ORF2p-driven Alu retrotransposition, and L1 retrotransposition

Identification of the conserved WD pair in the Cryptic sequence suggested that its absence from the EN289Δ construct analysed in Figure 8 could be contributing to the EN289Δ's inability to support BAR. We hypothesized that if the highly conserved WD pair is essential for ORF2p function, then its inclusion in the EN-containing ORF2p fragment would allow BAR despite the presence of a gap in the corresponding ORF2p primary sequence. We generated an EN314Δ construct designed to express a Gal4-tagged ORF2p fragment containing the WD (amino acids 288 and 289) and surrounding sequence (through amino acid 314) (Figure 12A). When coexpressed with the Δ348RTCys protein fragment, the combination of these two protein fragments introduced a 33 amino acid gap in the ORF2p sequence in the context of BAR. Western blot analysis with anti-Gal4 antibodies (Santa Cruz) detected a protein product of expected size in HeLa cells transiently transfected with the EN314Δ expression plasmid (Figure 12B). The BAR assay in HeLa cells demonstrated that the EN314Δ protein fragment supported BAR with the Δ348RTCys protein fragment at a level similar to the EN347Δ and Δ348RTCys protein fragment combination, despite the presence of a 33 amino acid gap relative to the full-length ORF2p sequence (Figure 12C).

To determine whether the WD pair in the EN-containing protein fragment and in the full-length ORF2p is required for function in BAR and conventional ORF2p-driven Alu mobilization, EN347Δ, ENZ490Δ, and ORF2 expression plasmids were modified using site-directed mutagenesis to express corresponding proteins with WD:AA mutations in the Cryptic sequence (Figure 13A). Mutant WD:AA and functional WD protein fragments and ORF2p were expressed at similar levels in HeLa cells as detected *via* western blot analysis with anti-Gal4 antibodies (Santa Cruz) (Figure 13B). Despite there being no noticeable change in protein expression between functional and mutant proteins, mutation of the WD in either the EN-containing protein fragments or full-length ORF2p abolished BAR and Alu mobilization in HeLa cells (Figure 13C). The same outcome *vis-à-vis* protein expression and BAR was observed with individual mutations (W288A, D289A) introduced within the WD pair in the ENZ490Δ protein fragment (Supplementary Figure S6B and S6C). The WD 288–289 AA mutation of the Cryptic WD pair also abolished Alu retrotransposition when driven by full-length L1 containing these ORF2 mutations (Supplementary Figure S6D). Introduction of the WD 288–289 AA mutations in the full-length L1 element severely compromised L1 retrotransposition in
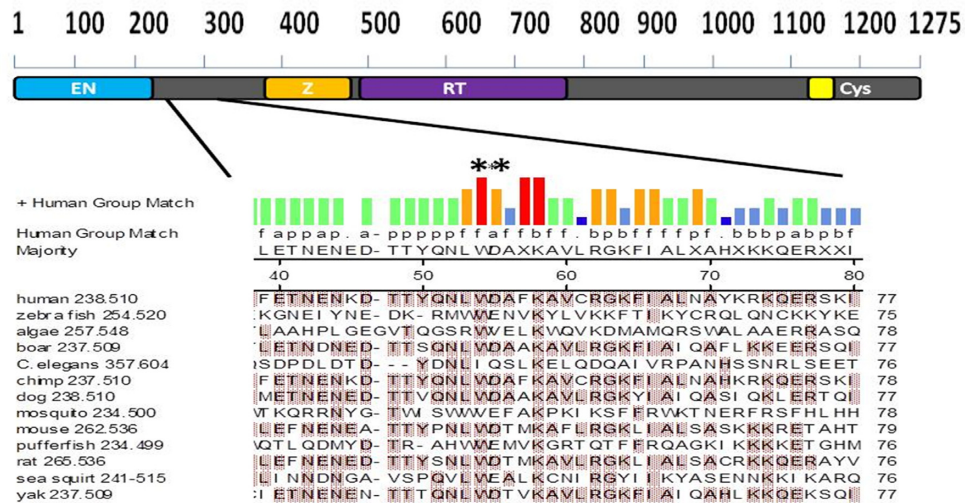
Figure 11. Alignment of the amino acid sequences (from C-terminal of EN domain to N-terminal of RT domain) from ORF2p and ORF2p-equivalent molecules reveals area of high conservation. The origin of the ORF2p and ORF2p-equivalent sequences used for alignment are indicated. Amino acids used shown to the right of each species. Above histogram represents strength of agreement of amino acid group to human amino acid sequence. Area of interest is denoted by double asterisks (WD). These amino acids correspond to human ORF2p amino acids 288 and 289.

HeLa cells (Figure 14). Further, the ability of the Cryptic ORF2 fragment to increase Alu retrotransposition driven by the RTCys-containing fragment is dependent upon the presence of the WD in Cryptic (Supplementary Figure S7). The ability of the Cryptic fragment to affect this increase in Alu retrotransposition is not due to an increase in steady-state protein levels of the RTCys-containing protein fragment (Supplementary Figure S8). Lastly, the abolishment of BAR upon mutation of the WD is not due to an increase in the cytotoxic potential of the EN-containing protein fragment (Supplementary Figures S9 and S10). We had previously reported that the addition of the Cryptic sequence to EN-containing ORF2p fragments caused a decrease in cytotoxic potential (8). Thus, we determined that Cryptic does not alter the cytotoxic potential of the cytotoxic EN-containing protein fragment (Supplementary Figure S11).

With this in mind, we also assessed the ability of the WD to effect the activity of the ORF2p RT domain using the previously described LEAP protocol (18,19) (Figure 15). Despite the equal expression of the ORF2p with functional WD and mutated WD, and the equal integrity of the RNA in the LEAP sample preparations, the ORF2p with mutated WD was unable to generate the expected cDNA product (Figure 15). These data establish a novel pair of amino acids in the ORF2p that are important for the Alu and L1 replication cycles through their effect on the ability of the ORF2p to generate cDNA.

To determine whether the WD is essential on one particular or both ORF2 protein fragments in BAR, we performed a BAR experiment similar to that described for the PIP YY mutation (Figure 6). We generated a longer RTCys-containing fragment (Δ263RTCys) that contained the WD (Figure 16A). Protein fragments containing functional and mutated WD were expressed at comparable levels (Figure 16B). While mutation of WD in both ORF2 protein fragments ablated BAR (Figure 16C: EN347Δ AA with Δ263RTCys AA, ENZ490Δ AA with Δ263RTCys

AA), we observed that as long as one fragment possessed a functional WD, BAR was supported (Figure 16C). However, mutation of the WD pair within the EN-containing fragments did reduce BAR efficiency when these EN-containing fragments were supplied with the RTCys-containing fragment with intact WD, while the reciprocal combination (mutant RTCys-containing fragment with intact EN-containing fragment) resulted in no decrease in BAR efficiency (Figure 16C).

DISCUSSION

The L1 ORF2p is a multifunctional protein critical to the retrotransposition process (1–3,6). However, how this protein operates in the mammalian cellular environment remains largely unknown. This is in part due to the fact that over 50% of the ORF2p currently has no known function associated with it (Figure 1). To date, most studies of the ORF2p have centred on the function of the two enzymatic domains—the endonuclease (EN) and reverse transcriptase (RT) domains (18,20,23,28–32). These domains have been studied using proteins purified from non-mammalian species or crudely purified gross cellular lysates (31,32). Structure/function studies of the EN have become possible with the resolution of the crystal structure of the minimal EN domain of the ORF2p (23). While the crystal structure of the L1 RT domain has not been solved yet, the ORF2p RT domain bears all the sequence hallmarks of an RT (24,33). Despite these advances, it remains unknown whether all of the ORF2p sequence is essential for retrotransposition, or if the unannotated portions of the L1 ORF2p sequence serve as simple linker sequences between the known functional domains and regions of the ORF2p. The limited size of the L1 genome and the complexity of its replication cycle support that there are important amino acids in the unannotated ORF2p regions that are essential for retrotransposition (Figures 7–16). Understanding the functional importance of these unannotated regions would
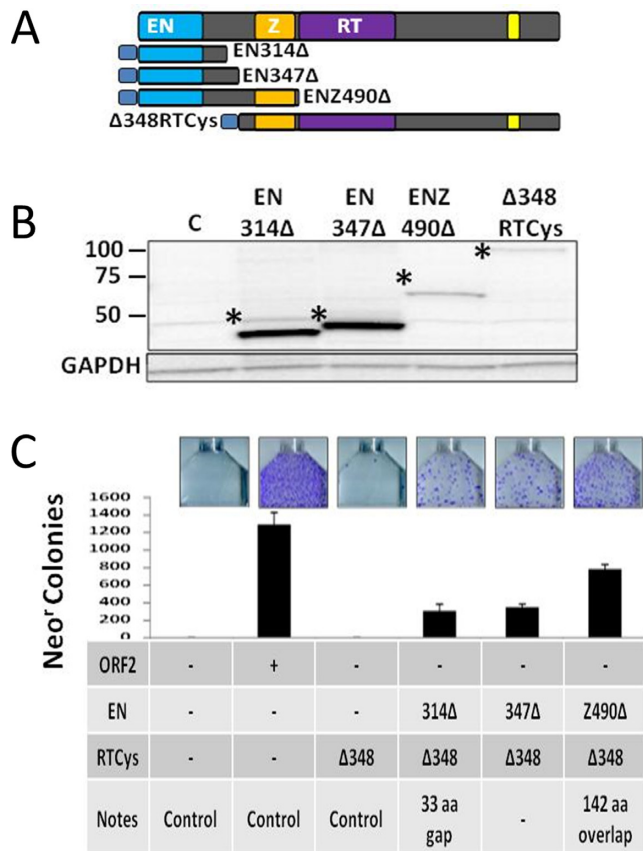
Figure 12. EN-containing fragment containing conserved amino acids between the EN and Z domain supports BAR with a gap in the ORF2p sequence. (**A**) Schematic representation of the Gal4-tagged (gray–blue square) ORF2 fragments generated from codon optimized ORF2 sequence. The EN314Δ fragment contains the EN domain and 75 amino acids between the EN and Z domains (amino acids 1–314). (**B**) Western blot analysis of ORF2p fragments with Gal4 DNA binding domain antibodies (Santa Cruz). Bands of expected sizes are denoted by black asterisks. Control lane (C) is total cell lysate from HeLa cells transfected with empty vector. Molecular weight markers are indicated on the left. GAPDH used as loading control. (**C**) BAR output in HeLa cells. Error bars denote standard deviation ($n = 3$). Representative flasks are shown above corresponding graph bars.

Figure 13. Conserved WD pair is essential for retrotransposition. (**A**) Schematic representation of the Gal4-tagged (grey-blue square) ORF2 fragments generated from codon optimized ORF2 sequence. EN-containing constructs were generated with nonfunctional WD (amino acids 288–289) with the approximate location denoted by white asterisks. (**B**) Western blot analysis of ORF2p and its fragments using Gal4 DNA binding domain antibodies (Santa Cruz). Bands of expected sizes are denoted by black asterisks. Control lane (C) is total cell lysate from HeLa cells transfected with empty vector. Molecular weight markers indicated at left. GAPDH used as loading control. (**C**) BAR output in HeLa cells. Error bars denote standard deviation ($n = 3$). Representative flasks are shown above corresponding graph bars.

significantly advance our understanding of retroelement biology in general, as well as the relevance of ORF2p DNA damage to human health.

Using BAR, we identified two new areas of importance to retrotransposition within the ORF2p sequence—the putative FF PCNA binding site (FF PCNA) and the WD pair of amino acids. Both are contained in a previously unstudied region of the molecule we have dubbed Cryptic (Figure 17). Previous work demonstrated the importance of the PCNA binding site present in the Z domain to the retrotransposition process (5). Protein motif alignment (Supplementary Figure S2) allowed us to identify a potential additional, putative PCNA binding domain (FF PCNA) in the Cryptic region of the ORF2p. Our data concerning the FF PCNA (Figure 7) suggest that one of the functions of the Cryptic sequence could be to serve as a docking station for proteins needed in the retrotransposition process. The possibility of the presence of multiple PCNA binding domains within the
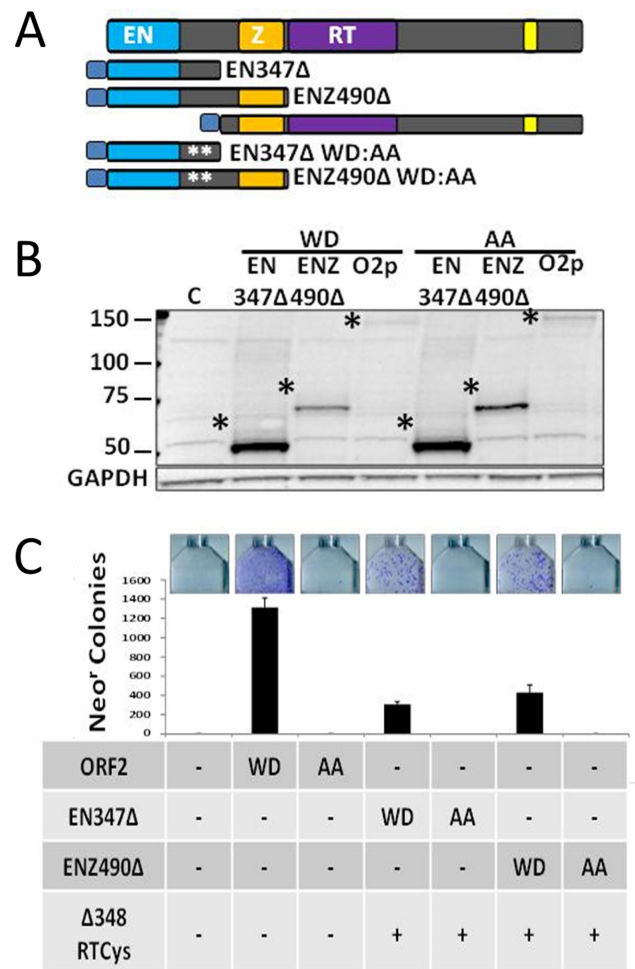
ORF2p sequence further supports the involvement of the PCNA trimer in coordinating the enzymatic functions of ORF2p during retrotransposition (5).

In addition to the discovery of the importance of the putative FF PCNA binding site to retrotransposition, BAR aided in the identification of the evolutionary conserved WD pair within the Cryptic sequence of ORF2p (Figure 11). We demonstrated that the WD to AA mutation abolished both BAR as well as Alu retrotransposition driven by the full-length ORF2p (Figure 13). The same WD mutation introduced into the full-length L1 element significantly ablated L1 retrotransposition (Figure 14). We also confirmed that both the W and D residues are essential for BAR, as independent mutation of either one of these two amino acids abolished BAR (Supplementary Figure S6). These amino
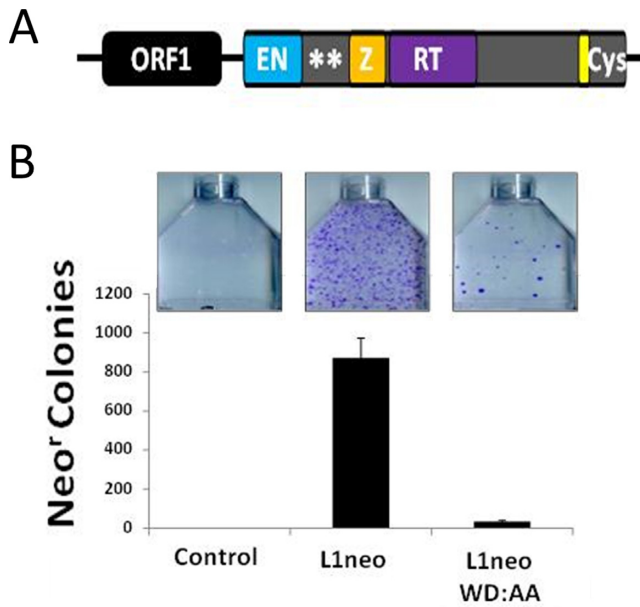
Figure 14. Conserved WD pair is important for L1 retrotransposition. (**A**) Schematic representation of the ORF2p molecule with location of WD denoted by white asterisks. The WD pair was mutated in the context of the full-length L1 in the codon-optimized L1 retrotransposition reporter construct. (**B**) Results of the L1 retrotransposition reporter assay with mutated WD. Mutation of the WD in the context of the full-length L1 significantly reduces the ability of L1 to retrotranspose. Error bars denote standard deviation ($n = 3$). Representative flasks are shown above corresponding graph bars.

acids appear to be important for cDNA synthesis (Figure 15). However, at what step in cDNA synthesis these amino acids exert their effect remains unknown. Our data suggest that the WD pair is either directly or indirectly influencing RT enzymatic activity or ORF2 protein binding to the parental RNA.

Previously, our lab has observed that when expressed in mammalian cells longer EN-containing fragments that include the Cryptic sequence are less cytotoxic than the minimal EN domain (8). The data shown here are consistent with this observation, as the EN fragments that contain the most Cryptic sequence exhibit the lowest amount of toxicity (Supplementary Figure S3). However, mutation of the WD in the context of tagged EN fragments does not appear to affect toxicity (Supplementary Figues S9 and S10). This suggests that other parts of the Cryptic region may be important for the observed decrease in toxicity, independent of or in combination with WD.

It was previously thought that the ORF2p could only function in retrotransposition as a single contiguous molecule. Our data demonstrate that this is not strictly the case for Alu (Figures 2 and 3). We demonstrated that the EN- and RT-containing protein fragments of ORF2p can support minimal Alu retrotransposition in human cells (Bipartite Alu Retrotransposition – BAR, Figure 2). However, the addition of the Gal4 DNA binding domain to the N-terminus of each fragment of the ORF2p significantly improved BAR output (Figure 3), allowing us to use BAR to interrogate the importance of ORF2p sequence with a more robust signal range than would be allowed using untagged
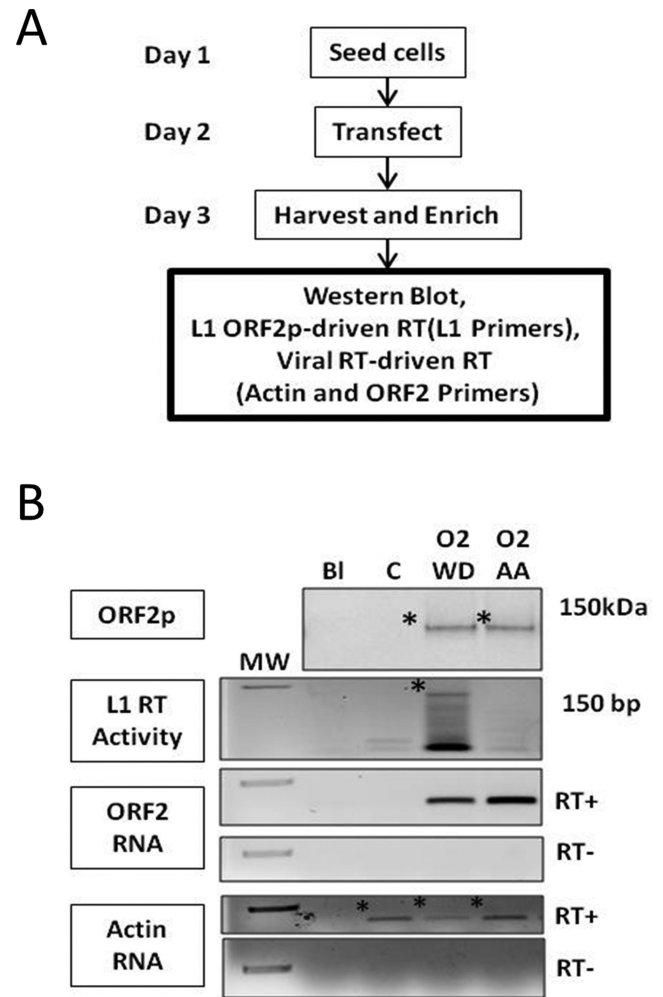

Figure 15. LEAP analysis of Gal-4 tagged ORF2p with functional and mutated WD. (**A**) Schematic of LEAP assay. Cells are seeded, transfected, and harvested for the purpose of enriching for ORF2p RNPs. The resulting enriched samples can be analysed for protein and RNA content. Using the RT activity of the ORF2p, a primer is used to add a linker sequence to the end of mRNA associated with the ORF2p. This linker-containing cDNA can then be amplified *via* PCR with a primer containing the linker sequence added in the RT step and a primer with sequence specific to the ORF2p sequence. Location of primers and PCR product are shown below. (**B**) Western blot, L1 RTPCR, and RTPCR analysis of LEAP samples. Protein levels between two ORF2p variants (O2 and O2 WD) shown to be equivalent using Gal4 antibodies. Blank (Bl) is loaded with loading buffer only and control (C) is LEAP prep on cells transfected with empty vector. L1 RTPCR using L1 ORF2p RT activity with primers described in (A) show a PCR band of expected molecule weight (MW) in the ORF2p LEAP sample with functional WD, but not in mutant WD ORF2p (O2 *vs.* O2 WD). Test of RNA integrity in LEAP preps were assessed using commercial RTPCR kit for Actin. Actin bands present in all three cellular LEAP samples in reverse transcriptase positive (RT+) reactions and not in control reactions (RT-).

ORF2p fragments. This Gal4-induced increase in BAR efficiency provides some potentially relevant insights into the ORF2p function.

It is known that the Gal4 DNA binding domain functions as a dimer (22). This suggests the possibility that EN and RT homodimers as well as EN/RT heterodimers may be formed upon coexpression (Supplementary Figure S13).
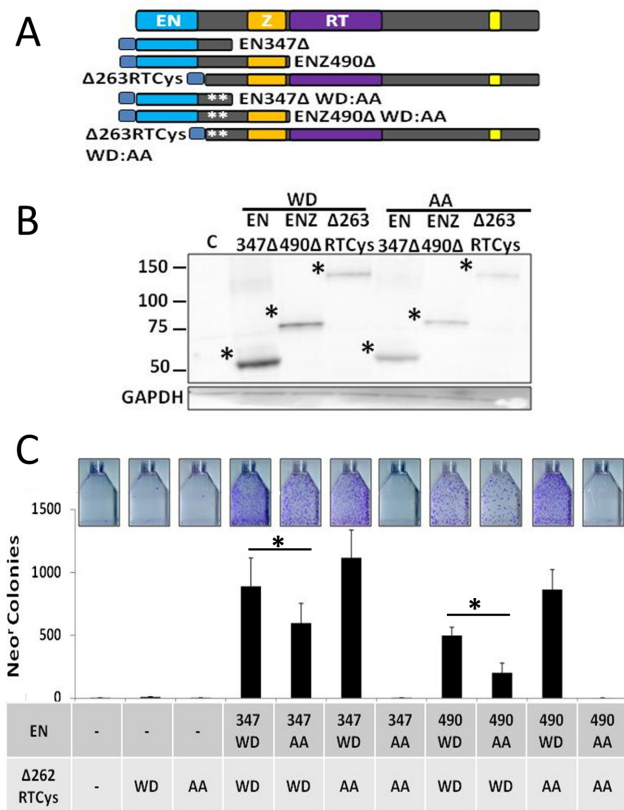
Figure 16. BAR analysis of longer RTCys-containing fragment with EN-containing fragments. (**A**) Schematic of Gal4-tagged (grey-blue square) ORF2 protein fragments generated from codon optimized DNA sequence used. For mutated WD (amino acids 288–289) fragments, approximate location denoted by white asterisks. (**B**) Western blot analysis of ORF2 protein fragments using Gal4 antibodies (Santa Cruz). Bands of expected sizes are denoted by asterisks. Control lane (C) is total cell lysate from HeLa cells transfected with empty vector. Molecular weight markers indicated at left. GAPDH used as loading control. (**C**) BAR output in HeLa cells. Error bars denote standard deviation ($n = 3$). Statistical significance assessed using Student's *t*-test and denoted by an asterisk. Representative flasks are shown above corresponding graph bars.
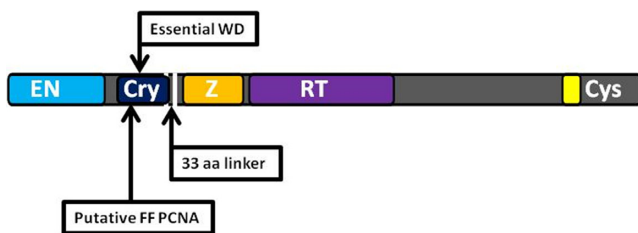

Figure 17. Cartoon of new areas of importance to retrotransposition discovered using BAR. The putative FF PCNA binding domain (FF PCNA) and WD are contained within the Cryptic sequence (Cry: dark blue). The 33 amino acid linker region dispensable for BAR is indicated by a white gap.

The dimerization of Gal4 supports that the N-termini of the EN- and RT-containing protein fragments in BAR may be brought into proximity to one another. While it is known that the Gal4 DNA binding domain does function as a dimer, we did not succeed in co-immunoprecipitating EN- and RT-containing fragments tagged with Gal4 (Supple-

mentary Figure S1). However, it has been reported that the ORF2p-equivalent molecule of the R2 retroelement functions as a dimer, and it has been theorized that the L1 ORF2p may dimerize as well (20,21). The R2 ORF protein is the *in vitro* model system used to study retrotransposition (34). Although the relative placement of the functional domains within the R2 protein is different than within the L1 ORF2p (35), many mechanistic aspects of R2 retrotransposition have been confirmed to apply to the L1 replication cycle. Our data demonstrate that the addition of the Gal4 tag to just the EN-containing ORF2p fragment improved BAR sixty-fold over the use of untagged EN- and RT-containing fragments (Figure 3). It is known that many DNA endonucleases function as dimers (Reviewed in (36,37)), and our findings support the possibility that the ORF2p EN-containing fragments may do the same in BAR. Many retroviral reverse transcriptases also function as dimers (Reviewed in (25,38)). Consistent with this possibility, our results demonstrate that the addition of the Gal4 tag to the N-terminus of the RT-containing fragment further improved BAR efficiency. It would be interesting to know if dimerization of the untagged L1 ORF2p is important for retrotransposition, as our data support that EN and RT dimerization may be important for facilitating BAR (Supplementary Figure S13). It is, however, of note that this is but one explanation for why Gal4 fusion to ORF2p fragments improves the BAR signal.

Another discovery made using the BAR system was relevant to the reported *cis* preference of the L1 ORF2p for its parental mRNA as a retrotransposition substrate (18). While the importance of *cis* preference for L1 propagation and the defence against amplification of host RNA by the retrotransposition machinery is well appreciated, the specific mechanism of this preference is currently unknown. Our data demonstrate that the Δ348RTCys protein fragment may be the rate-limiting protein fragment in BAR (Figures 4 and 5). While the EN-containing ORF2 fragments produced from wild-type or codon optimized DNA sequences supported BAR at equivalent efficiencies when coexpressed with the RT-containing fragment made from codon optimized sequence, their combination with the Δ348RTCys fragment derived from wild-type ORF2 DNA sequence produced minimal BAR signal (Figures 4 and 5). These data suggest that the RTCys portion of ORF2p molecule used in BAR may be important for the *cis* preference of the full-length protein. The Δ348RTCys fragment contains protein sequence that has been implicated in nucleic acid binding, which may be responsible for the cis-preference of the L1 ORF2p (7).

Developing BAR has contributed in many ways to advancing our understanding of the ORF2p function in retrotransposition. While it cannot be assumed that discoveries made using BAR will always translate into the full-length ORF2p or L1, our data demonstrate that the method has significant utility as a first-line tool for beginning to understand the importance of the unannotated region of the molecule. The discovery of the significance of the Cryptic region to mobilization and the characterization of the WD mutation added to a finite number of individual amino acids that are absolutely essential for ORF2p function. Previously, such amino acids were restricted to the known, an-

notated domains of the molecule (3,6). The putative FF PCNA and WD pair are the first characterized amino acid sequences shown to be essential to retrotransposition in the area of the ORF2p molecule between the EN and Z domains. We anticipate that in the future BAR will continue aiding in the identification of important, previously unannotated areas of interest within the ORF2p molecule. Even though the exact mechanism of protein fragment interaction in BAR remains to be identified and we could not co-immunoprecipitate these fragments (Supplementary Figure S1), our results support the existence of potential associations between the ORF2p fragments that could be important for retrotransposition driven by the full-length, linear molecule. This interaction could be a direct association between the two fragments or indirect tethering through an RNA or cellular protein intermediate. Identifying the mechanism of these putative interactions in BAR could have direct relevance to the protein:protein or protein:RNA interactions taking place during the Alu and L1 replication cycles, and will be a focus of future studies.

In conclusion, we identified amino acids important to retrotransposition (Figure 17. The ORF2p is currently thought to be involved in a variety of human disease processes. The expression of the ORF2p in cultured human cells leads to the generation of DNA double strand breaks (DSB) and retrotransposition (8,10,39). There is now considerable data linking L1 activity with cancer (40–42) and other age-related conditions typically associated with genomic instability (Reviewed in (43,44)). An understanding of the role of the unannotated ORF2p sequence in the retrotransposition process and function of the ORF2p in a biologically relevant manner would help to identify cellular factors controlling L1-induced genomic instability, as well as specific pathological conditions potentially associated with an increase in L1-mediated DNA damage. However, how the ORF2p is specifically implicated in these processes is currently unknown. It is possible that the Cryptic region of ORF2p is important to the impact of ORF2p and L1 (as well as Alu) on human health. Future work will focus on the mechanistic importance of these areas, as well as using BAR to identify additional such areas within the Cryptic region.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENT

FUNDING

REFERENCES

1. Feng,Q., Moran,J.V., Kazazian,H.H. and Boeke,J.D. (1996) Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. *Cell*, **87**, 905–916.
2. Mathias,S.L., Scott,A.F., Kazazian,H.H., Boeke,J.D. and Gabriel,A. (1991) Reverse transcriptase encoded by a human transposable element. *Science*, **254**, 1808–1810.
3. Clements,A.P. and Singer,M.F. (1998) The human LINE-1 reverse transcriptase:effect of deletions outside the common reverse transcriptase domain. *Nucleic Acids Res.*, **26**, 3528–3535.
4. Jamburuthugoda,V.K. and Eickbush,T.H. (2014) Identification of RNA binding motifs in the R2 retrotransposon-encoded reverse transcriptase. *Nucleic Acids Res.*, **42**, 8405–8415.
5. Taylor,M.S., Lacava,J., Mita,P., Molloy,K.R., Huang,C.R., Li,D., Adney,E.M., Jiang,H., Burns,K.H., Chait,B.T. *et al.* (2013) Affinity proteomics reveals human host factors implicated in discrete stages of LINE-1 retrotransposition. *Cell*, **155**, 1034–1048.
6. Moran,J.V., Holmes,S.E., Naas,T.P., DeBerardinis,R.J., Boeke,J.D. and Kazazian,H.H. (1996) High frequency retrotransposition in cultured mammalian cells. *Cell*, **87**, 917–927.
7. Piskareva,O., Ernst,C., Higgins,N. and Schmatchenko,V. (2013) The carboxy-terminal segment of the human LINE-1 ORF2 protein is involved in RNA binding. *FEBS Open Biol.*, **3**, 433–437.
8. Kines,K.J., Sokolowski,M., deHaro,D.L., Christian,C.M. and Belancio,V.P. (2014) Potential for genomic instability associated with retrotranspositionally-incompetent L1 loci. *Nucleic Acids Res.*, **42**, 10488–10502.
9. Wallace,N., Wagstaff,B.J., Deininger,P.L. and Roy-Engel,A.M. (2008) LINE-1 ORF1 protein enhances Alu SINE retrotransposition. *Gene*, **419**, 1–6.
10. Wallace,N.A., Belancio,V.P. and Deininger,P.L. (2008) L1 mobile element expression causes multiple types of toxicity. *Gene*, **419**, 75–81.
11. Belancio,V.P., Roy-Engel,A.M., Pochampally,R.R. and Deininger,P. (2010) Somatic expression of LINE-1 elements in human tissues. *Nucleic Acids Res.*, **38**, 3909–3922.
12. Morrish,T.A., Gilbert,N., Myers,J.S., Vincent,B.J., Stamato,T.D., Taccioli,G.E., Batzer,M.A. and Moran,J.V. (2002) DNA repair mediated by endonuclease-independent LINE-1 retrotransposition. *Nat. Genet.*, **31**, 159–165.
13. Wagstaff,B.J., Kroutter,E.N., Derbes,R.S., Belancio,V.P. and Roy-Engel,A.M. (2013) Molecular reconstruction of extinct LINE-1 elements and their interaction with nonautonomous elements. *Mol. Biol. Evol.*, **30**, 88–99.
14. Perepelitsa-Belancio,V. and Deininger,P. (2003) RNA truncation by premature polyadenylation attenuates human mobile element activity. *Nat. Genet.*, **35**, 363–366.
15. Dewannieux,M., Esnault,C. and Heidmann,T. (2003) LINE-mediated retrotransposition of marked Alu sequences. *Nat. Genet.*, **35**, 41–48.
16. deHaro,D., Kines,K.J., Sokolowski,M., Dauchy,R.T., Streva,V.A., Hill,S.M., Hanifin,J.P., Brainard,G.C., Blask,D.E. and Belancio,V.P. (2014) Regulation of L1 expression and retrotransposition by melatonin and its receptor: implications for cancer risk associated with light exposure at night. *Nucleic Acids Res.*, **42**, 7694–7707.
17. Sokolowski,M., Deharo,D., Christian,C.M., Kines,K.J. and Belancio,V.P. (2013) Characterization of L1 ORF1p Self-Interaction and Cellular Localization Using a Mammalian Two-Hybrid System. *PLoS One*, **8**, e82021.
18. Kulpa,D.A. and Moran,J.V. (2006) Cis-preferential LINE-1 reverse transcriptase activity in ribonucleoprotein particles. *Nat. Struct. Mol. Biol.*, **13**, 655–660.
19. Wagstaff,B.J., Barnerssoi,M. and Roy-Engel,A.M. (2011) Evolutionary conservation of the functional modularity of primate and murine LINE-1 elements. *PLoS One*, **6**, e19672.
20. Christensen,S.M. and Eickbush,T.H. (2005) R2 target-primed reverse transcription: ordered cleavage and polymerization steps by protein subunits asymmetrically bound to the target DNA. *Mol. Cell. Biol.*, **25**, 6617–6628.
21. Christensen,S.M., Bibillo,A. and Eickbush,T.H. (2005) Role of the Bombyx mori R2 element N-terminal domain in the target-primed reverse transcription (TPRT) reaction. *Nucleic Acids Res.*, **33**, 6461–6468.

22. Carey,M., Kakidani,H., Leatherwood,J., Mostashari,F. and Ptashne,M. (1989) An amino-terminal fragment of GAL4 binds DNA as a dimer. *J. Mol. Biol.*, **209**, 423–432.

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

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

25. Herschhorn,A. and Hizi,A. (2010) Retroviral reverse transcriptases. *Cell. Mol. Life Sci.*, **67**, 2717–2747.

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

27. Permyakov,S.E., Permyakov,E.A. and Uversky,V.N. (2015) Intrinsically disordered caldesmon binds calmodulin via the 'buttons on a string' mechanism. *PeerJ*, **3**, e1265.

28. Barré-Sinoussi,F., Chermann,J.C., Rey,F., Nugeyre,M.T., Chamaret,S., Gruest,J., Dauguet,C., Axler-Blin,C., Vézinet-Brun,F., Rouzioux,C. *et al.* (1983) Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science*, **220**, 868–871.

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

30. Dhellin,O., Maestre,J. and Heidmann,T. (1997) Functional differences between the human LINE retrotransposon and retroviral reverse transcriptases for in vivo mRNA reverse transcription. *EMBO J.*, **16**, 6590–6602.

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

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

33. Eickbush,T.H. and Jamburuthugoda,V.K. (2008) The diversity of retrotransposons and the properties of their reverse transcriptases. *Virus Res.*, **134**, 221–234.

34. Xiong,Y.E. and Eickbush,T.H. (1988) Functional expression of a sequence-specific endonuclease encoded by the retrotransposon R2Bm. *Cell*, **55**, 235–246.

35. Burke,W.D., Malik,H.S., Jones,J.P. and Eickbush,T.H. (1999) The domain structure and retrotransposition mechanism of R2 elements are conserved throughout arthropods. *Mol. Biol. Evol.*, **16**, 502–511.

36. Wilce,J., Vivian,J. and Wilce,M. (2012) Oligonucleotide binding proteins: the occurrence of dimer and multimer formation. *Adv. Exp. Med. Biol.*, **747**, 91–104.

37. Pingoud,A., Wilson,G.G. and Wende,W. (2014) Type II restriction endonucleases—a historical perspective and more. *Nucleic Acids Res.*, **42**, 7489–7527.

38. Hizi,A. and Herschhorn,A. (2008) Retroviral reverse transcriptases (other than those of HIV-1 and murine leukemia virus): a comparison of their molecular and biochemical properties. *Virus Res.*, **134**, 203–220.

39. Gasior,S.L., Wakeman,T.P., Xu,B. and Deininger,P.L. (2006) The human LINE-1 retrotransposon creates DNA double-strand breaks. *J. Mol. Biol.*, **357**, 1383–1393.

40. Solyom,S., Ewing,A.D., Rahrmann,E.P., Doucet,T., Nelson,H.H., Burns,M.B., Harris,R.S., Sigmon,D.F., Casella,A., Erlanger,B. *et al.* (2012) Extensive somatic L1 retrotransposition in colorectal tumors. *Genome Res.*, **22**, 2328–2338.

41. Rodić,N., Sharma,R., Zampella,J., Dai,L., Taylor,M.S., Hruban,R.H., Iacobuzio-Donahue,C.A., Maitra,A., Torbenson,M.S., Goggins,M. *et al.* (2014) Long interspersed element-1 protein expression Is a Hallmark of many human cancers. *Am. J. Pathol.*, **184**, 1280–1286.

42. Doucet-O'Hare,T.T., Rodić,N., Sharma,R., Darbari,I., Abril,G., Choi,J.A., Young Ahn,J., Cheng,Y., Anders,R.A., Burns,K.H. *et al.* (2015) LINE-1 expression and retrotransposition in Barrett's esophagus and esophageal carcinoma. *Proc. Natl. Acad. Sci. U.S.A.*, **112**, E4894–E4900.

43. Belancio,V.P., Roy-Engel,A.M. and Deininger,P.L. (2010) All y'all need to know 'bout retroelements in cancer. *Semin. Cancer Biol.*, **20**, 200–210.

44. Sturm,Á., Ivics,Z. and Vellai,T. (2015) The mechanism of ageing: primary role of transposable elements in genome disintegration. *Cell. Mol. Life Sci.*, **72**, 1839–1847.