

METHODS AND COMPOSITIONS FOR GENOMIC INTEGRATION

CROSS REFERENCE

[0001] This application claims the benefit of priority to U.S. Provisional Application No. 63/187,117, filed on May 11, 2021, U.S. Provisional Application No. 63/254,791, filed on October 12, 2021, and U.S. Provisional Application No. 63/274,907, filed on November 2, 2021, each of which is hereby incorporated by reference in its entirety.

BACKGROUND

[0002] Cell therapy is a rapidly developing field for addressing difficult to treat diseases, such as cancer, persistent infections and certain diseases that are refractory to other forms of treatment. Cell therapy often utilizes cells that are engineered ex vivo and administered to an organism to correct deficiencies within the body. An effective and reliable system for manipulation of a cell's genome is crucial, in the sense that when the engineered cell is administered into an organism, it functions optimally and with prolonged efficacy. Likewise, reliable mechanisms of genetic manipulation form the cornerstone in the success of gene therapy. However, severe deficiencies exist in methods for delivering nucleic acid cargo (e.g., large cargo) in a therapeutically safe and effective manner. Viral delivery mechanisms are frequently used to deliver large nucleic acid cargo in a cell but are tied to safety issues and cannot be used to express the cargo in some cell types. Additionally, subjecting a cell to repeated gene manipulation can affect cell health, induce alterations of cell cycle and render the cell unsuitable for therapeutic use. Advancements are continually sought in the area for efficacious delivery and stabilization of an exogenously introduced genetic material for therapeutic purposes.

SUMMARY

[0003] Provided herein is a pharmaceutical composition comprising a therapeutically effective amount of one or more polynucleic acids, or at least one vector encoding the one or more polynucleic acids, the one or more polynucleic acids comprising: a mobile genetic element comprising a sequence encoding a polypeptide; and an insert sequence, wherein the insert sequence comprises a sequence that is a reverse complement of a sequence encoding an exogenous therapeutic polypeptide, wherein the polypeptide encoded by the sequence of the mobile genetic element promotes integration of the insert sequence into a genome of a cell; and wherein the pharmaceutical composition is substantially non-immunogenic to a human subject.

[0003a] In one aspect, provided herein is a of modifying a population of human cells, comprising contacting the population with one or more polynucleic acids comprising:

- (a) a mobile genetic element, wherein the mobile genetic element comprises a sequence encoding LINE-1 polypeptides comprising:
 - (i). a human ORF1p or a functional fragment thereof, and
 - (ii). a human ORF2p or a functional fragment thereof, wherein the human ORF2p or functional fragment thereof comprises two NLSs that are the same or different;

(b) an insert sequence, wherein the insert sequence is a reverse complement of a sequence comprising a gene or a fragment thereof, wherein the polypeptides encoded by the sequence of the mobile genetic element promotes integration of the sequence comprising the gene or the fragment thereof into a genome of a cell of the population of human cells via target-primed reverse transcription (TPRT),

wherein the one or more polynucleic acids comprise one or more mRNA molecules, and wherein the sequence comprising the gene or fragment thereof integrates into the genome at a locus that is not a ribosomal RNA locus.

[0003b] In another aspect, provided herein is a composition comprising one or more polynucleic acids comprising one or more mRNA molecules, comprising:

- (a) a 5'UTR sequence;
- (b) a mobile genetic element, wherein the mobile genetic element comprises a sequence encoding LINE-1 polypeptides comprising:
 - (i). a human ORF1p or a functional fragment thereof, and
 - (ii). a human ORF2p or a functional fragment thereof, wherein the human ORF2p or functional fragment thereof comprises 2 NLSs that are the same or different;
- (c) insert sequence, wherein the insert sequence is a reverse complement of a sequence comprising a gene or a fragment thereof; and
- (d) a 3'UTR sequence downstream of the insert sequence.

[0003c] In another aspect, provided herein is a method of expressing an exogenous human therapeutic polypeptide encoded by an exogenous sequence from a site-specifically integrated genomic DNA sequence of a target human cell, the method comprising:

- (a) contacting a composition to the target human cell, the composition comprising an RNA molecule encoding a Cas nickase and one or more RNA molecules, wherein the target human cell uptakes the one or more RNA molecules, and wherein the one or more RNA molecules comprise a first RNA molecule comprising an RNA sequence encoding a Cas nickase and a second RNA molecule comprising an RNA sequence that comprises:
 - (i). a sequence that is a reverse complement of a DNA sequence that encodes the exogenous human therapeutic polypeptide; and
 - (ii). a mobile genetic element, wherein the mobile genetic element comprises an RNA sequence encoding a polypeptide with target-primed reverse transcription (TPRT) activity, wherein the polypeptide with TPRT activity comprises an endonuclease with a mutation that abrogates endonuclease activity of the endonuclease domain;
- (b) translating the RNA sequence encoding the polypeptide with TPRT activity and translating the RNA sequence encoding the Cas nickase;

- (c) nicking the genomic DNA site of the target human cell via the Cas nickase translated in (b);
- (d) reverse transcribing the sequence that is the reverse complement of the DNA sequence that encodes the human therapeutic polypeptide via the TPRT activity of the polypeptide with TPRT activity translated in (b), thereby producing a DNA sequence encoding the exogenous human therapeutic polypeptide;
- (e) integrating the DNA sequence encoding the exogenous human therapeutic polypeptide produced in (d) into a genomic DNA site of the target human cell; and
- (f) expressing the exogenous human therapeutic polypeptide in the target human cell, wherein the exogenous human therapeutic polypeptide is expressed from the DNA sequence integrated into the genomic DNA of the target human cell in (e).

[0003d] In another aspect, provided herein is a composition comprising:

- (i). an RNA molecule encoding a Cas nickase and
- (ii). one or more RNA molecules, wherein the target human cell uptakes the one or more RNA molecules, and wherein the one or more RNA molecules comprise a first RNA molecule comprising an RNA sequence encoding a Cas nickase and a second RNA molecule comprising an RNA sequence that comprises:
 - (a) a sequence that is a reverse complement of a DNA sequence that encodes the exogenous human therapeutic polypeptide; and
 - (b) a mobile genetic element, wherein the mobile genetic element comprises an RNA sequence encoding a polypeptide with target-primed reverse transcription (TPRT) activity, wherein the polypeptide with TPRT activity comprises an endonuclease with a mutation that abrogates endonuclease activity of the endonuclease domain.

[0004] In some embodiments, the polypeptide encoded by the sequence of the mobile genetic element comprises one or more long interspersed nuclear element (LINE) polypeptides, wherein the one or more LINE polypeptides comprises: human ORF1p or a functional fragment thereof, and human ORF2p or a functional fragment thereof.

[0005] In some embodiments, the insert sequence stably integrates and/or is retrotransposed into the genome of a human cell.

that is a reverse complement of a 5' UTR sequence is downstream of the sequence that is a reverse complement of a sequence encoding an exogenous therapeutic polypeptide, (iii) the sequence that is a reverse complement of a 3' UTR sequence is upstream of the sequence that is a reverse complement of a sequence encoding an exogenous therapeutic polypeptide, and (iv) the sequence that is a reverse complement of a poly A sequence is upstream of the sequence that is a reverse complement of a 3' UTR sequence and downstream of the sequence of the mobile genetic encoding a polypeptide. In some embodiments, the promoter sequence of the first expression cassette is different from the promoter sequence of the second expression cassette. In some embodiments, the one or more LINE polypeptides comprises a first LINE polypeptide comprising the human ORF1p or functional fragment thereof and a second LINE polypeptide comprising the human ORF2p or functional fragment thereof, wherein the first LINE polypeptide and the second LINE polypeptide are translated from different open reading frames (ORFs). In some embodiments, the one or more polynucleic acids comprises a first polynucleic acid molecule encoding the human ORF1p or functional fragment thereof and a second polynucleic acid molecule encoding the human ORF2p or functional fragment thereof. In some embodiments, the one or more polynucleic acids comprises a 5' UTR sequence and a 3' UTR sequence, wherein the 5' UTR comprises a 5' UTR from LINE-1 or a sequence with at least 80% sequence identity to

ACUCCUCCCCAUCCUCUCCCUCUGUCCCUCUGUCCCUCUGACCCUGCACUGUCCCAGCACC (SEQ ID NO: 51); and/or the 3' UTR comprises a 3' UTR from LINE-1 or a sequence with at least 80% sequence identity to

CAGGACACAGCCUUGGAUCAGGACAGAGACUUGGGGGCCAUCCUGCCCCUCCAACCCGACAUGUGUACCUCAGCUUUUCCCUCACUUGCAUCAAUAAAGCUUCUGUGUUUGGAACAG (SEQ ID NO: 52). In some embodiments, the sequence encoding the exogenous therapeutic polypeptide does not comprise introns. In some embodiments, the polypeptide encoded by the sequence of the mobile genetic element comprises a C-terminal nuclear localization signal (NLS), an N-terminal NLS or both. In some embodiments, the sequence encoding the exogenous polypeptide is not in frame with a sequence encoding the ORF1p or functional fragment thereof and/or is not in frame with a sequence encoding the ORF2p or functional fragment thereof. In some embodiments, the one or more polynucleic acids comprises a sequence encoding a nuclease domain, a nuclease domain that is not derived from ORF2p, a megaTAL nuclease domain, a TALEN domain, a Cas9 domain, a Cas6 domain, a Cas7 domain, a Cas8 domain, a zinc finger binding domain from an R2 retroelement, or a DNA binding domain that binds to repeat sequences. In some embodiments, the one or more polynucleic acids comprises a sequence encoding the nuclease domain, wherein the nuclease domain does not have nuclease activity or comprises a mutation that reduces activity of the nuclease domain compared to the nuclease domain without the mutation. In some embodiments, the ORF2p or functional fragment thereof lacks endonuclease activity or comprises a mutation selected from the

specification, the specification is intended to supersede and/or take precedence over any such contradictory material.

BRIEF DESCRIPTION OF THE DRAWINGS

[0015] The novel features of the invention are set forth with particularity in the appended claims. A better understanding of the features and advantages of the present invention will be obtained by reference to the following detailed description that sets forth illustrative embodiments, in which the principles of the invention are utilized, and the accompanying drawings (also "FIG." herein), of which:

[0016] FIG. 1A illustrates a general mechanism of action of retrotransposons. (I) is a schematic representing the overall lifecycle of an autonomous retrotransposon. (II) LINE-1 retrotransposon comprises LINE-1 elements, which encode two proteins ORF1p and ORF2p that are expressed as mRNAs. The bicistronic mRNA is translated into the two proteins, and when ORF2p is translated by a read-through event by the ribosome, it binds the 3' end of its own mRNA through the poly A tail (III). ORF2p cleaves at a consensus sequence TAAAA, where the poly A at the 3' end of the mRNA hybridizes and primes the reverse transcriptase activity of the ORF2 protein. The protein reverse-transcribes the mRNA back into DNA leading to an insertion of the LINE-1 sequence back into a new location in the genome (IV).

[0017] FIG. 1B is an illustration of a schematic diagram of an mRNA construct that comprises a genetic payload (left) that can be designed for integration into the genome (right).

[0018] FIG. 1C illustrates various exemplary designs for integrating an mRNA encoding a transgene into the genome of a cell. GFP shown here in a box is an exemplary transgene.

[0019] FIG. 1D illustrates various exemplary designs for integrating an mRNA encoding a transgene into the genome of a cell. GFP shown here in a box is an exemplary transgene.

[0020] FIG. 1E is an illustration a schematic of the LINE-1 retrotransposition cycle showing the mechanism of action of the LINE transposons and introduction of a transgene cargo into a retrotransposon cite. LINE-1 retrotransposons are genomic sequences that encode for two proteins, ORF1 and ORF2. These elements are transcribed and translated into proteins that form an RNA-protein complex with the LINE-1 mRNA, ORF1 trimers, and ORF2, a reverse-transcriptase endonuclease. This complex translocates back into the nuclease where it cleaves DNA at a 5'-TTTT N-3' motif and is primed for reverse-transcription of the LINE-1 RNA by the ORF2 protein by making an RNA-DNA hybrid with the poly A tail of the mRNA and the resected cleaved DNA. Reverse-transcription of the LINE-1 into cDNA leads to a new LINE-1 integration event.

[0021] FIG. 2A illustrates three exemplary designs for expressing an exemplary transgene GFP by stably incorporating the sequence encoding GFP using the constructs. Expected GFP expression levels at 72 hours are shown on the right side.

[0022] FIG. 2B illustrates three exemplary designs for expressing an exemplary transgene GFP by stably incorporating the sequence encoding RFP, RFP and GFP or ORF2p and GFP using the constructs. Expected GFP and RFP expression levels at 72 hours are shown on the right side.

[0023] FIG. 3A illustrates an exemplary diagram of conventional circRNA structure and formation.

[0024] FIG. 3B illustrates two views of an exemplary RL-GAAA tectoRNA motif designs. FIG. 3B discloses SEQ ID NOS 111-112, respectively, in order of appearance.

[0025] FIG. 3C illustrates exemplary structures of chip-flow piece RNAs as platforms for testing potential tectoRNA.

[0026] FIG. 4A illustrates an exemplary schematic showing ORF2p binding to an ORF2 poly A region.

[0027] FIG. 4B illustrates an exemplary schematic showing how a fusion of ORF2p with an MS2 RNA binding domain binds to an MS2 binding RNA sequence in the 3'UTR of an mRNA encoding the ORF2 an increase specificity.

[0028] FIG. 4C illustrates exemplary designs of retrotransposon systems for stably integrating a nucleic acid into the genome of a cell at specific sites. The upper panel shows a design using an ORF2p-MegaTAL DNA binding domain fusion where the DNA binding and endonuclease activity of ORF2p is mutated to be inactive. The middle panel shows a chimeric ORF2p where the endonuclease domain has been replaced with a high specificity and high-fidelity nuclease domain of another protein. The lower panel shows a fusion of a DNA binding domain of a heterologous protein with ORF2p such that the fusion protein binds to ORF2 binding site as well additional DNA sequences in the vicinity of the ORF2 site.

[0029] FIG. 5 illustrates exemplary constructs (I) – (X) for integrating an mRNA encoding a transgene into the genome of a cell.

[0030] FIG. 6A illustrates an exemplary construct with a sequence encoding ORF1p for integrating an mRNA encoding a transgene into the genome of a cell.

[0031] FIG. 6B illustrates an exemplary construct without a sequence encoding ORF1p for integrating an mRNA encoding a transgene into the genome of a cell.

[0032] FIG. 7A illustrates exemplary methods of improving mRNA half-life by inhibiting degradation by 5'-3' exonucleases, such as XRN1, or 3'-5' exosomal degradation, by introducing structures corresponding to a G-quadruplex, or, a pseudoknot (SEQ ID NOS 113) in the 5'UTR; and/or xrRNAs, a triplex motifs (SEQ ID NOS 74, 54, and 53, respectively, in order of appearance) and/or a non-A nucleotide residues in the 3'UTR.

[0033] FIG. 7B illustrates an exemplary schematic of a myeloid cell expressing a transgene encoding a chimeric receptor that binds a cancer cell and induces anti-cancer activity.

[0034] FIG. 7C shows expected results of introducing bulk or purified RNA encoding a chimeric receptor that binds a cancer cell as described in FIG. 7B on increased and prolonged expression of the chimeric receptors.

[0035] FIG. 8A shows an exemplary plasmid design and expected LINE-1 mRNA transcript with a cargo nucleic acid sequence. The plasmid has a LINE-1 sequence (comprising ORF1 and ORF2 protein encoding sequences) and a cargo sequence which is a nucleic acid sequence encoding GFP, where the coding sequence of GFP is interrupted with an intron. The GFP is not expressed until the sequence is integrated in the genome and the intron is spliced.

[0036] FIG. 8B shows exemplary results showing successful integration of the mRNA transcript encoded by the plasmid shown in FIG. 8A and expression of GFP relative to mock-transfected cells (fold

increase in mean fluorescence intensity of GFP positive cells is shown). Mock transfected cells were transfected by the vector lacking the GFP cargo sequence.

[0037] FIG. 8C shows exemplary flow cytometry results from the results shown in **FIG. 8B**.

[00153] In some embodiments, the one or more polynucleic acids comprises a 5' UTR sequence and a 3' UTR sequence, wherein (a) the 5' UTR comprises a 5' UTR from LINE-1 or a sequence with at least 80% sequence identity to

ACUCCUCCCCAUCCUCUCCUCUGUCCUCCUGUCCUCCUGACCCUGCACUGUCCCAGCACC
(SEQ ID NO: 51); and/or (b) the 3' UTR comprises a 3' UTR from LINE-1 or a sequence with at least 80% sequence identity to

CAGGACACAGCCUUGGAUCAGGACAGAGACUUGGGGGCCAUCCUGCCCCUCCAACCCGAC
AUGUGUACCUCAGCUUUUCCCUCACUUGCAUCAAUAAAGCUUCUGUGUUUGGAACAG
(SEQ ID NO: 52).

[00154] In some embodiments, the sequence encoding the exogenous therapeutic polypeptide does not comprise introns.

[00155] In some embodiments, the polypeptide encoded by the sequence of the mobile genetic element comprises a C-terminal nuclear localization signal (NLS), an N-terminal NLS or both.

[00156] In some embodiments, the sequence encoding the exogenous polypeptide is not in frame with a sequence encoding the ORF1p or functional fragment thereof and/or is not in frame with a sequence encoding the ORF2p or functional fragment thereof.

[00157] In some embodiments, the one or more polynucleic acids comprises a sequence encoding a nuclease domain, a nuclease domain that is not derived from ORF2p, a megaTAL nuclease domain, a TALEN domain, a Cas9 domain, a Cas6 domain, a Cas7 domain, a Cas8 domain, a zinc finger binding domain from an R2 retroelement, or a DNA binding domain that binds to repeat sequences.

[00158] In some embodiments, the one or more polynucleic acids comprises a sequence encoding the nuclease domain, wherein the nuclease domain does not have nuclease activity or comprises a mutation that reduces activity of the nuclease domain compared to the nuclease domain without the mutation.

[00159] In some embodiments, the ORF2p or functional fragment thereof lacks endonuclease activity or comprises a mutation selected from the group consisting of S228P and Y1180A, and/or wherein the ORF1p or functional fragment comprises a K3R mutation.

[00160] In some embodiments, the insert sequence comprises a sequence that is a reverse complement of a sequence encoding two or more exogenous therapeutic polypeptides.

[00161] In some embodiments, the one or more polynucleic acids comprises one or more polyribonucleic acids, wherein the exogenous therapeutic polypeptide is a receptor selected from the group consisting of a chimeric antigen receptor (CAR) and a T cell receptor (TCR), and wherein the pharmaceutical composition is formulated for systemic administration to a human subject.

[00162] In some embodiments, the one or more polynucleic acids (i) are formulated in a nanoparticle selected from the group consisting of a lipid nanoparticle and a polymeric nanoparticle; and/or (ii) comprises one or more polynucleic acids selected from the group consisting of glycosylated RNAs, circular RNAs and self-replicating RNAs.

[00163] Also provided herein is a method of treating a disease or condition in a human subject in need thereof comprising administering a pharmaceutical composition described herein to the human subject.

that exhibits a transferred transposon end sequence as follows: 5' AGATGTGTATAAGAGACAG 3' (SEQ ID NO: 55), and a non-transferred strand that exhibits a "non-transferred transposon end sequence" as follows: (5' CTGTCTCTTATACACATCT 3' (SEQ ID NO: 56). The 3'-end of a transferred strand is joined or transferred to target DNA in an *in vitro* transposition reaction. The non-transferred strand, which exhibits a transposon end sequence that is complementary to the transferred transposon end sequence, is not joined or transferred to the target DNA in an *in vitro* transposition reaction.

[00286] In some embodiments, the transferred strand and non-transferred strand are covalently joined. For example, in some embodiments, the transferred and non-transferred strand sequences are provided on a single oligonucleotide, e.g., in a hairpin configuration. As such, although the free end of the non-transferred strand is not joined to the target DNA directly by the transposition reaction, the non-transferred strand becomes attached to the DNA fragment indirectly, because the non-transferred strand is linked to the transferred strand by the loop of the hairpin structure. As used herein an "cleavage domain" refers to a nucleic acid sequence that is susceptible to cleavage by an agent, e.g., an enzyme.

[00287] A "restriction site domain" means a tag domain that exhibits a sequence for the purpose of facilitating cleavage using a restriction endonuclease. For example, in some embodiments, the restriction site domain is used to generate di-tagged linear ssDNA fragments. In some embodiments, the restriction site domain is used to generate a compatible double-stranded 5'-end in the tag domain so that this end can be ligated to another DNA molecule using a template-dependent DNA ligase. In some embodiments, the restriction site domain in the tag exhibits the sequence of a restriction site that is present only rarely, if at all, in the target DNA (e.g., a restriction site for a rare-cutting restriction endonuclease such as NotI or AscI).

[00288] As used herein, the term "recombinant nucleic acid molecule" refers to a recombinant DNA molecule or a recombinant RNA molecule. A recombinant nucleic acid molecule is any nucleic acid molecule containing joined nucleic acid molecules from different original sources and not naturally attached together. Recombinant RNA molecules include RNA molecules transcribed from recombinant DNA molecules. A recombinant nucleic acid may be synthesized in the laboratory. A recombinant nucleic acid can be prepared by using recombinant DNA technology by using enzymatic modification of DNA, such as enzymatic restriction digestion, ligation, and DNA cloning. A recombinant DNA may be transcribed *in vitro*, to generate a messenger RNA (mRNA), the recombinant mRNA may be isolated, purified and used to transfect a cell. A recombinant nucleic acid may encode a protein or a polypeptide. A recombinant nucleic acid, under suitable conditions, can be incorporated into a living cell, and can be expressed inside the living cell. As used herein, "expression" of a nucleic acid usually refers to transcription and/or translation of the nucleic acid. The product of a nucleic acid expression is usually a protein but can also be an mRNA. Detection of an mRNA encoded by a recombinant nucleic acid in a cell that has incorporated the recombinant nucleic acid, is considered positive proof that the nucleic acid is "expressed" in the cell. The process of inserting or incorporating a nucleic acid into a cell can be via transformation, transfection or transduction. Transformation is the process of uptake of foreign nucleic acid by a bacterial cell. This process is adapted for propagation of plasmid DNA, protein production, and other applications.

[00361] In some embodiments, the retrotransposon may comprise a SINE or LINE element. In some embodiments, the retrotransposon comprises a SINE or LINE stem loop structure, such as an Alu element.

[00362] In some embodiments, the retrotransposon is a LINE-1 (L1) retrotransposon. In some embodiments, the retrotransposon is human LINE-1. Human LINE-1 sequences are abundant in the human genome. There are approximately 13,224 total human L1s, of which 480 are active, which make up about 3.6%. Therefore, human L1 proteins are well tolerated and non-immunogenic in humans. Moreover, a tight regulation of random transposition in human ensures that random transposase activity will not be triggered by introduction of the L1 system as described herein. In addition, the retrotransposable constructs designed herein may comprise targeted and specific incorporation of the insert sequence. In some embodiments, the retrotransposable genetic element may comprise designs intended to overcome the silencing machinery actively prevalent in human cells, while being careful that random integration resulting in genomic instability is not initiated.

[00363] Accordingly, the retrotransposable constructs may comprise a sequence encoding a human LINE-1 ORF1 protein; and a human LINE-1 ORF2 protein. In some embodiments, the construct comprises a nucleic acid sequence encoding an ORF1p protein with at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to
MGKKQNRKTGNSKTQSASPPPKERSSSPATEQSWMENDFDELREEGFRRSNYSELREDIQTKGK
EVENFEKNLEECITRITNTEKCLKELMELKTKARELREECRSLRSRCDQLEERVSAMEDEMNE
KREGKFREKRIKRNEQSLQEIWDYVKRPNLRLIGVPESDVENGTKLENTLQDIIQENFPNLARQA
NVQIQEIQRTPQRYSSRRATPRHIIVRFTKVEMKEKMLRAAREKGRVTLKGKPIRLTVDLAETL
QARREWGPFIENILKEKNFQPRISYPAKLSFISEGEIKYFIDKQMLRDFVTTRPALKELLKEALNME
RNNRYQPLQNHAKM (SEQ ID NO: 57). In some embodiments, the construct comprises a nucleic acid sequence with at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to
atgggcaagaagcaaaatcgcaagacggggaattccaagacacaatccgctagcccaccacctaagagcgttctagctccctgctactgagcagtc
ctggatggaaaacgacttcgatgaactccgggaagagggatttagcgatccaactattcagaactccggaagatatccagacaaaggggaaggaa
gtcgagaatttcgagaagaacctcgaggagtgcattcacccgtatcacaacactgagaaatgtctcaagaactcatggaacttaagacaaaagccag
ggagcttcgagaggagtgtcggagtctgagatccaggtgtgaccagctcgaggagcgcgtgagcgcgatggaagacgagatgaacgagatgaaa
gagagggcaaatcaggagaagcgcattaagaggaacgaacagagtctgcaggagatttgggattacgtcaagaggcctaactgcggtgatcgg
cgtccccgagagcgacgtagaaaacgggactaaactggagaatacactcaagacatcattcaagaaaatttccaaacctggctcggaagctaattgt
gcaaatcaagagatccaacgcacaccccgcggtatagctctcggcgtgccaccctaggcatattatcgtgcgtttactaagggtggagatgaaaga
gaagatgctgcgagccgctcgggaaaagggaagggtgactttgaagggcaaacctattcggtgacggttgaccttagcgccgagacactccaggca
cgccgggaatggggcccatcttaatatctgaaggagaagaactccagccacgaatctctaccctgcaaagttgagtttatctccgaggggtgagat
taagtatttcacgataaacagatgctgcgagacttcgtgacaactcgccagctctcaaggaaactgctcaagaggctcttaatatggagcgcaataata
gatatcaacccttgacagaaccacgcaaagatgtga (SEQ ID NO: 58).

[00364] In some embodiments, the construct comprises a nucleic acid sequence encoding an ORF2p protein with at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to MTGNSHITILTLNINGLNSAIKRHLASWIKSQDPSVCCIQETHLTCRDTHRLKIKGWRKIYQANG KQKKAGVAILVSDKTDFKPTKIKRDKEGHYIMVKGSIQQEELTILNIYAPNTGAPRFIKQVLSDLQR DLDSHTLIMGDFNTPLSTLDRSTRQKVNKDTQELNSALHQADLIDIYRTLHPKSTEYTFFSAPHHTY SKIDHIVGSKALLSKCKRTEIITNYLSHSAIKLELRIKNLTQSRSTTWKLNNLLNDYWVHNEMK AEIKMFFETNENKDDTTYQNLWDAFKAVCRGKFIALNAYKRRQERSKIDTLTSQLKELEKQEQTHS KASRRQEITKIRAELEIETQKTLQKINESRSWFFERINKIDRPLARLIKKKREKNQIDTIKNDKGDIT TDPTEIQTIREYYKHL YANKLENLEEMDTFLDITYTLPRLNQEEVESLNRPIITGSEIVAIINSLPTKKS PGPDGFTAIFYQRYMEELVPFLKLFQSIEKEGILPNSFYEASIIIPKPGRDITTKENFRPISLMNID AKILNKILANRIQQHIKKLIHHDQVGFIPGMQGWFNIRKSINVIQHINRAKDKNHMIISIDAEKAFDK IQQPFMLKTLNKLIGDGTIFYKIIRAIYDKPTANIILNGQKLEAFPLKTGTRQGCPLSPLLFNIVLEVLA RAIHQEKEIKGIQLGKEEVKLSLFADDMIVYLENPIVSAQNLLKLISNFSKVSGYKINVQKSQAFLY TNNRQTESQIMGELPFVIASKRIKYLGIQLTRDVKDLFKENYKPLLKEIKEDTNKWKNIPCSWVGRI NIVKMAILPKVIYRFNAIPIKLPMTFFTELEKTTLKFIWNQKRARIAKSILSQKNKAGGITLPDFKLY YKATVTKTAWYWYQNRDIDQWNRTEPSEIMPHIYNYLIFDKPEKNKQWKGDSL FNKWCWENWL AICRKLKLDPFLTPYTKINSRWIKDLNVKPKTIKTLEENLGITIQDIGVGKDFMSKTPKAMATKDKI DKWDLIKLKSFC TAKETTIRVNRQPTTWEKIFATYSSDKGLISRIYNELKQIYKKKTNNPIKKWAK DMNRHFSKEDIYAAKKHMKKCSSSLAIREMQIKTTMRYHLTPVRMAIHKSGNNRCWRGCGEIGT LLHCWWDCCLVQPLWKS VWRFLRDLELEIPFDPAPILLGIYPNEYKSCCYKDTCTRMFIAALFTIA KTNWQPKCPTMIDWIKKMWHIYTMEYYAAIKNDEFISFVGTWMKLETIILSKLSQEQKTKHRIFSL IGGN (SEQ ID NO: 59). In some embodiments, the construct comprises a nucleic acid sequence with at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to atgaccggtctaaactacatatcaccatccttacacttaacattaacggcctcaactcagctatcaagcgccatcggtggccagctggatcaaatacagg atccaagcggttgatccaagagacccacctgacctgtagagatactcaccgctcaagatcaaggatggcgaaagatttatcaggcgaacggtaag cagaagaaagccggagtcgcaattctggtctcagacaagacggattcaagcccacaaaattaagcgtgataagggaaggtcactatattatggtgaaagg cagcatacagcaggaagaacttaccatattgaacatctacgcgccaacacccggcgacacctcgtttatcaaacaggtcctgtccgatctgcagcgagatct ggattctcatagcttgattatgggtgattcaatacaccattgagcacctggatcgagcaccaggcaaaaggttaataaagacacgcaagagctcaatag cgcactgcatcaggcagatctcattgatattatcgactcttcacctaagagtaccgagtagacacattctcagcgccccacatcatactcaaagatcga tcatatcgctggctcaaaggctctgctgtcaagtgcaagcgcacagagataattacaaattacctgtcagatcatagcgcatcaagctcgagctgagaat caagaacctgaccagagccggagtagcacttggaagcttaataacctgctgctcaacgattattgggtccacaatgagatgaaggcagagattaaaatgtt ctctgaacaataatgagaataaggatactacatcaaaaccttgggatgccttaaggccgtctgcagaggcaaggtcatcgccctcaacgcctataaaga aaacaagagagatctaagatcgatactctcacctctcagctgaaggagtgagaaacaggaacagaccactccaaggcgtaagacggcaggagatc acaagattcgcgccgagtgtaagagatcgaaacccaaaagactcttcagaaaattaacgagtcctcgtagttggttctt

cgagcggattaataagatagacagacctctggcacgactgattaagaagaagcgcgaaaagaaccagattgataccatcaagaacgacaagggcgacat
cactactgacccgaccgagatccagaccactattcgggagtattataagcatttgatgtaacaagcttgagaacctggaagagatggacacttttctggata
cctatactctgccacggcttaatacaagaggaagtcgagtcctcaaccgccaattacaggaagcgagattgtggccataattaacctctccgacaaga
aatctcctgggtccggacgggtttacagctgagttttatcaacggtatatggaagagctgtaccgtttctgctcaagctcttcagctctatagaaaaggaaggca
tcttgcceaattcctctacgaagcttctataatacttattccaaaccaggacgcgataccacaaaagaaggaaaactccggcccattagtctcatgaatatcg
acgctaaaatattgaacaagattctcgccaacagaatccaacaacatattaagaaattgatacatcacgaccaggtgggtttatacctggcatgcagggtg
gtttaacatccggaagagtattaacgtcattcaacacattaatagagctaaggataagaatcatatgatcatctctatagacgcggaagggcattcgataaga
ttcagcagccatttatgctcaagactctgaacaaactcggaatcgacggaacatattttaagattattcgcgcaatttacgataagccgactgtaacattatcct
taacggccaaaagctcgaggccttccgctcaagactggaacccgccaaggtgtccctctccccgctttgtttaattgtactcgaggtgctggctagg
gctattcgtcaagagaagagattaagggatagacgtcgggaaggaagaggtcaagcttctgttcgccgatgatattgtgtacctggagaatccta
ttgtgtctgctcagaaccttctaaacttatttcaacttttagcaaggtcagcggctataagattaacgtccagaaatctcaggccttctgtacacaaataatcga
cagaccgaatcccagataatgggtgagcttccgtttgtcatagccagcaaaaggataaagtatctcggaatccagctgacacgagacgttaagattgttta
aggaaaattacaagcctctctgaagagattaaggaagataactaataagtggaagaatatccctgttcattgggtggcagaatcaacatagtgaagatgg
caatacttctaagtgatataatcgtttaacgccatcccaattaaactgcctatgaccttcttacggagctcgagaaaacaaccttaatttatatggaatcaa
aagagagcaagaatagcgaagtcctcttgagccagaagaataaggccggtgggattacttgcctgatttaagttgtattataaagccacagtaactaaga
cagcctggattggtatcagaatagagacatcgaccagtgaatcgaccgaacctatcagagataatgccccacatctataattaccttatattcgataagcc
agaaaagaataaacagtggtggcgaagacagcctcttcaacaagtggtgtgggagaattggctggccatatgccggaactcaagctcgacccctttctta
cacctacactaaaatcaacagtaggtggatcaaggacttgaatgtcaagccaaagactataaagacactggaagagaatcttgggatcacaatacaagat
ataggcgtcggcaagattttatgtcaagacgccaaggccatggccactaaggataagattgataagtgggaccttattaagctcaaaagcttctgtactg
ccaaggagaccagatcagagttataggcagccactacatgggaaaagatttccgacctattcatcagataaggggttgataagcagaatataaacg
agctgaagcagatctacaagaagaaaacgaataatcccatcaagaagtgggcaaaagatatgaacaggcattttagcaaagaggatatctacgcccgcaa
gaagcatatgaagaagtgtagtcaagcttggccattcgtgagatgcagattaagacgacctgcgataccaccttccccagtgaggatggcaattatcaa
gaaatctggcaataatagatgttggcgggctgtggcgagattggcaccctgctccattgctggtgggattgcaagctggtgcagccgctttggaaatcagt
ctggcgctttctgagggacctcgagcttgagattcccttcgatcccgcaattcccttgctcggaatctatcctaacgaatacaagagctgtgttacaaggatac
gtgtacccggatgttcacgcggccttgttacgatatagtaagacgtggaatcagcctaagtgccccacaatgatcgttggtatcaagaaaatgtggcatattt
ataccatggagtattacgcagcaattaagaatgacgaatttatttctctgttgggacctggatgaagctggagactattattctgagcaagctgtctcaggagc
aaaagacaaaagcatagaattctctctcattggtgtaactaa (SEQ ID NO: 60).

[00365] In some embodiments, the construct comprises a nucleic acid sequence encoding an ORF2p protein with at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to
MVGITYISIITLVNGLNAPTKRHRLAEWIQKQDPYICCLQETHFRPRDYLRLKVRGWKKIFHANG
NQKKAGVAILISKIDFKIKNVTRDKEGHYIMIQGSIQEEDITIINIYAPNIGAPQYIRQLLTAIKEEID
SNTIIVGDFNTSLTPMDRSSMKINKETALNDTIDQIDLIDIYRTFHPKTADYTFSSAHGTFSRIDH
ILGHKSSLSKFKKIEIISIFSDHNAMRLEMNHREKNVKKTNWRLNNTLLNNQEITEEIKQEIKKYL
ETNDNENTTTQNLWDAKAVLRGKFIAIQAYLKKQEKSQVNNLTLHLKKLEKEEQTKPKVSRRK
EIIKIRAEINEIETKKTIKINKTKSWFFEKINKIDKPLARLIKKKRERTQINKIRNEKGEVTTDTAEIQ
NILRDYYKQLYANKMDNLEEMDKFLERYNLPRLNQEETENINRPITSNEIETVIKNL

PTNKSPGPDGFTGEFYQTFREELTPILLKLFQKIAEEGTLPSNFYEATITLIPKPKD DTTKKENYRPI
 SLMNIDAKILNKILANRIQQHIKRIIHHDDQVGFIPGMQGGFFNIRKSINVIHHINKLKKKNHMIISIDA
 EKAFDKIQHPFMIKTLQKVGIEGTYLNIKAIYDKPTANIILNGEKLKAFPLRSGTRQGCPLSPLLF
 NIVLEVLATAIREEKEIKGIQIGKEEVKLSLFADDMILYIENPKTATRKLLELINEYGKVAGYKINA
 QKSLAFLYTNDKSEREIMETLPFTIATKRIKYLGINLPKETKDLYAENYKTLMKEIKDDTNRWR
 DIPCSWIGRINIVKMSILPKAIYRFNAIPIKLPMAFFTELEQIILKFVWRHKRPRIAKAVLRQKNGA
 GGIRLPDFRLYYKATVIKTIWYWHKNRNIDQWNKIESPEINPRTYGQLIYDKGGKDIQWRKDSLF
 NKWCWENWTATCKRMKLEYSLTPYTKINSKWIRDLNIRLDTIKLLEENIGRTLFDINHSKIFFDPP
 PRVMEIKTKINKWDLMLKLSFCTAKETINKTKRQPSEWEKIFANESTDKGLISKIYKQLIQLNIKE
 TNTPIQKWAEDLNRHFSKEDIQTATKHMKRCSTSLIREMQIKTTMRYHLTPVRMGIIRKSTNNK
 CWRGCGEKGTLLHCWWECKLIQPLWRTIWRFLKKLKIELPYDPAIPLLGIIYPEKTVIQKDTCTR
 MFIAALFTIARSWKQPKCPSTDEWIKKMWYIYTMEYSSAIKRNEIGSFLETWMDLETVIQSEVSQ
 KEKNKYRILTHICGTWKNGTDEPVC RTEIETQM (SEQ ID NO: 61). In some embodiments, the
 construct comprises a nucleic acid sequence with at least 80%, at least 81%, at least 82%, at least 83%, at
 least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%,
 at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%,
 or 100% sequence identity to

atgggtcataggaacatacatatcgataattacctaactgaatggattaaatgccccaaccaaaagacatagactggctgaatggatacaaaaacaag
 acccatatatatgctgtctacaagagaccacttcagacctagggacacatacagactgaaagtgaggggatggaaaaagatattccatgcaaatggaa
 atcaaaagaaagctggagtagctatactcatatcagataaaatagactttaaaataaagaatgttacaagagacaaggaaggacactacataatgatcca
 gggatcaatccaagaagaagatataacaattataatataatgcaccaacataggagcacctcaatacataaggcaactgctaacagctataaaagag
 gaaatcgacagtaacacaataatagtgggggactttaacacctcacttacccaatggacagatcatccaaatgaaaataaataaggaaacagaagctt
 taaatgacacaatagaccagatagatttaattgatataataggacattccatccaaaacagcagattacacgttcttcaagtgcgcacggaacattctc
 caggtagatcacatcttgggtcacaaatcaagcctcagtaaatttaagaaaattgaaatcatatcaagcatctttctgaccacaacgctatgagattagaa
 atgaatcacagggaaaaaacgtaaaaaagacaaacacatggaggctaaacaatacgttactaaataaccaagagatcactgaagaatcaaacagga
 aataaaaaatacctagagacaaatgacaatgaaaacacgacgacccaaacctatgggatgcagcaaaagcgggttaagagggaagtttatagctat
 acaagcctacctaagaacaagaataatctcaagtaacaatctaacttacctaaagaactagagaaagaacaacaaaacccaaagtta
 gcagaaggaaagaatcataagatcagagcagaaataatgaaatagaacaaagaaacaatagcagaatcaataaaactaaaagtggttcttg
 agaagataacaaaattgataagccattagccagactcatcaagaaaaaggaggagactcaaatcaataaaatcagaatgaaaaaggagaagtt
 acaacagacaccgcagaatacaaaacatcctaagagactactacaagcaactttatgccaataaaatggacaacctggaagaatggacaaattctta
 gaaaggtataacctccaagactgaaccaggaagaacagaaaatcaacagaccaatcacaagtaatgaaatgaaactgtgattaaaaatcttcaa
 caaacaagaagtcaggaccagatggcttcacaggtgaattctatcaaacatttagagaagagctaacacccatccttctcaaaactcttcaaaaaattgca
 gaagaaggaaactccaaactcattctatgaggccaccatcacctgataccaaaaccagacaaagacactacaaaaaagaaattacagaccaat
 atcactgatgaatatagatgcaaaaatctcaacaaaatactagcaaacagaatccaacaacacattaaaaggatcataccacgatcaagtgggattta
 tcccagggatgcaaggattcttcaatatacgcaaatcaatcaatgtgataccatattaacaaattgaagaagaaaaaccatatgatcatctcaatagatg
 cagaaaaagcttttgacaaaattcaacaccatttatgataaaaactctccagaaagtgggcatagagggaacctacctcaacataataaaggccatatat
 gacaaacccacagcaaacatcattctcaatgggtgaaaaactgaaagcatttccttaagatcaggaacgagacaaggatgtccactctcaccactattatt
 caacatagttctggaagtctagccacggcaatcagagaagaaaaagaataaaaggaaatacaaatggaaaagaagaagtaaaactgtcactgtttgc

ggatgacatgatactatacatagagaatcctaaaactgccaccagaaaactgctagagctaattaatgaatatggtaaagttgcaggttacaaaattaatgc
acagaaatctcttgcaatcctatacactaatgatgaaaaatctgaaagagaaattatggaaactcccattaccattgcaacaaaaagaataaaataccta
ggaataaacctacctaaggagacaaaagacctgtatgcagaaaactataagacactgatgaaagaaattaaagatgataccaacagatggagagatata
ccatgttcttgattggaagaatcaacattgtgaaaatgagtatactacccaaagcaatctacagattcaatgcaatccctatcaaattaccaatggcattttt
acggagctagaacaaatcatcttaaaattgtatggagacacaaaagaccccgaaatgccaaagcagctctgaggcaaaaaaatggagctggaggaat
cagactccctgacttcagactatactacaaagctacagtaatacaagacaatatggtactggcacaaaaacagaaacatagatcaatggacaagatagaa
agcccagagattaacccacgcacctatggtcaactaatctatgacaaaggaggcaagatatacaatggagaaaagacagtctcttaataagtgtgtct
gggaaaactggacagccacatgtaaaagaatgaaattagaatactccctaacaccatacacaaaaataaactcaaatggattagagacctaataataag
actggacactataaaactcttagaggaaaacataggaagaacactcttgacataaatcacagcaagatcttttcgatccacctcctagagtaattggaat
aaaaacaaaaataaacaagtgggacctaatgaaactcaaaacttttgacagcaaaaggaaaccataaacaagacgaaaagacaacccctcagaatggg
agaaaatatttgcaaatgaatacaacggacaaaggattaatctccaaatatataaacagctcattcagctcaatatcaaagaacaaacaccccaatcaa
aatgggcagaagacctaaatagacatttctccaaagaagacatacagacggccacgaagcacatgaaaagatgctcaacatcactaattattagagaa
atgcaaatcaaaactacaatgaggtatcacctcactctgttagaattgggcatcatcagaaaaatctacaaacaacaatgtggagagggtgtggagaaa
agggaaacctctgactgttggtgggaatgtaattgatacagccactatggagaacaatatggaggtccttaaaaaactaaaaatagaattaccatag
accagcaatcccactactgggcatataccagagaaaaccgtaattcaaaaagacacatgcacccgaatgttcattgcagcactattacaatagccag
gtcatggaagcaacctaatgcccacgacagacgaatggataaagaagatgtgttacatatatacaatggaatattactagccataaaaagggaacga
aattgggtcatttttagagacgtggatggatctagagactgtcatacagagtgaagtaagtcagaaagagaaaaacaaatatcgtatattaacgcataatg
tggaaacctggaaaaatgttacagatgaaccggtctgcaggacagaaattgagacacaaatgtaa (SEQ ID NO: 62).

[00366] In some embodiments, the construct comprises a nucleic acid sequence encoding a nuclear localization sequence with at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to PAAKRVKLD (SEQ ID NO: 63). In some embodiments, the nuclear localization sequence is fused to the ORF2p sequence. In some embodiments, the construct comprises a nucleic acid sequence encoding a flag tag having the sequence DYKDDDDK (SEQ ID NO: 64). In some embodiments, the flag tag is fused to the ORF2p sequence. In some embodiments, the flag tag is fused to the nuclear localization sequence.

[00367] In some embodiments, the construct comprises a nucleic acid sequence encoding an MS2 coat protein with at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to ASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEV PKGAWRSYLN MELTIPIFATNSDCELVKAMQGLLKDGNIPIPSAIAANSGIYAMASNFTQFVLVD NGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEV PKGAWRSYLN MELTIPIFATNSDCELVKAMQGLLKDGNIPIPSAIAANSGIY (SEQ ID NO: 65). In some embodiments, the MS2 coat protein sequence is fused to the ORF2p sequence.

[00368] In some embodiments, the transgene may comprise a flanking sequence which comprises an Alu ORF2p recognition sequence.

85%, about 80%, about 75%, about 70%, about 65%, about 60%, or about 50% of the ORF2 endonuclease domain is deleted.

[00397] In some embodiments, an additional nuclease domain is inserted into the ORF2 protein sequence. In some embodiments, ORF2 endonuclease domain is deleted, and is replaced with an endonuclease domain from a heterologous protein. In some embodiments, the ORF2 endonuclease is partially deleted and replaced with an endonuclease domain from a heterologous protein. The endonuclease domain from a heterologous protein may be a mega-TAL nuclease domain. The endonuclease domain from a heterologous protein may be a TALENs. The endonuclease domain from a heterologous protein may be a Cas9 loaded with a specific gRNA for a locus.

[00398] In some embodiments, the endonuclease is an endonuclease that has (i) a specific target on the genome and (ii) it creates a 5'-P and a 3'-OH terminus at the cleavage site.

[00399] In some embodiments, the additional endonuclease domain from a heterologous protein is an endonuclease domain from a related retrotransposon.

[00400] In some embodiments, the endonuclease domain from a heterologous protein may comprise a bacterial endonuclease engineered for targeting a specific site. In some embodiments, the endonuclease domain from a heterologous protein may comprise a domain of a homing endonuclease or a fragment thereof. In some embodiments, the endonuclease is a homing endonuclease. In some embodiments, the homing endonuclease is an engineered LAGLIDADG homing endonucleases (LHEs) ("LAGLIDADG" disclosed as SEQ ID NO: 66) or a fragment thereof. In some embodiments, additional endonucleases may be a restriction endonuclease, Cre, Cas TAL or fragments thereof. In some embodiments, the endonuclease may comprise a Group II intron encoded protein (ribozyme) or a fragment thereof.

[00401] An engineered or modified L1-ORF2p as discussed in the preceding paragraphs, that is endowed with specific DNA targeting capability due to the additional/heterologous endonuclease is expected to be highly advantageous in driving targeted stable integration of a transgene into the genome. The engineered L1-ORF2p can generate much reduced off-target effects when expressed in a cell than using a native, non-engineered L1-ORF2p. In some embodiments, the engineered L1-ORF2p generates no off-target effect.

[00402] In some embodiments, the engineered or modified L1-ORF2p targets a recognition site that is other than the usual $(Py)_n\downarrow(Pu)_n$ site. In some embodiments, engineered L1-ORF2p targets a recognition site that comprises the $(Py)_n\downarrow(Pu)_n$ site, for example, TTTT/AA site, such as a hybrid target site. In some embodiments, the engineered L1-ORF2p targets a recognition site having at least one nucleotide in addition to the conventional L1-ORF2 $(Py)_n\downarrow(Pu)_n$ site, for example TTTT/AAG, or TTTT/AAC, or TTTT/AAT, TTTT/AAA, GTTTT/AA, CTTTT/AA, ATTTT/AA, or TTTTT/AA. In some embodiments, the engineered L1-ORF2p targets a recognition site that is in addition to the conventional L1-ORF2p $(Py)_n\downarrow(Pu)_n$ site. In some embodiments, the engineered L1-ORF2p targets a recognition site that is other than to the conventional L1-ORF2p $(Py)_n\downarrow(Pu)_n$ site. In some embodiments, the engineered L1-ORF2p targets a recognition site that is 4, 5, 6, 7, 8, 9, 10 or more nucleotides long. In some embodiments, the engineered or modified L1-ORF2p recognition site may be 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 or more nucleotides.

(CAS 912444-00-9); Camptothecin (CPT) (CAS 7689-03-4); Irinotecan (CAS 100286-90-6); Topotecan (Hycamtin® GlaxoSmithKline) (CAS 123948-87-8); NSC 19630 (CAS 72835-26-8); NSC 617145 (CAS 203115-63-3); ML216 (CAS 1430213-30-1); 6-hydroxyDL- dopa (CAS 21373-30-8); D-103; D-G23; DIDS (CAS 67483-13-0); B02 (CAS 1290541-46-6); RI-1 (CAS 415713-60-9); RI-2 (CAS 1417162-36-7); Streptonigrin (SN) (CAS 3930-19-6).

III. Nucleic Acid Cargo:

A. Transgene

[00449] In one aspect the transgene or noncoding sequence that is the heterologous nucleic acid sequence to be inserted within the genome of a cell is delivered as an mRNA. The mRNA may comprise greater than about 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10,000 bases. In some embodiments, the mRNA may be more than 10,000 bases long. In some embodiments, the mRNA may be about 11,000 bases long. In some embodiments, the mRNA may be about 12,000 bases long. In some embodiments, the mRNA comprises a transgene sequence that encodes a fusion protein. In some embodiments, the nucleic acid is delivered as a plasmid.

[00450] In some embodiments, the nucleic acid is delivered in the cell by transfection. In some embodiments, the nucleic acid is delivered in the cell by electroporation. In some embodiments, the transfection or electroporation is repeated more than once to enhance incorporation of the nucleic acid into the cell.

[00451] Contemplated herein are retrotransposon mediated stable integration of a recombinant nucleic acid encoding a phagocytic or tethering receptor (PR) fusion protein (CFP). In some embodiments, the CFPs comprise: a PR subunit comprising: a transmembrane domain, and an intracellular domain comprising an intracellular signaling domain; and an extracellular domain comprising an antigen binding domain specific to an antigen of a target cell; wherein the transmembrane domain and the extracellular domain are operatively linked.

[00452] In some embodiments, the nucleic acid comprises a sequence encoding a chimeric fusion protein (CFP), the CFP comprising an extracellular domain comprising a CD5 binding domain, and a transmembrane domain operatively linked to the extracellular domain. In some embodiments, the CD5 binding domain is a CD5 binding protein, such as an antigen binding fragment of an antibody, a Fab fragment, an scFv domain or an sdAb domain. In some embodiments, wherein the CD5 binding domain comprises an scFv comprising (i) a variable heavy chain (VH) sequence with at least 90% sequence identity to EIQLVQSGGGLVKPGGSVRISCAASGYTFTNYGMNWVRQAPGKGLEWMGWINTHTG EPTYADSFKGRFTFSLDDSKNTAYLQINSLRAEDTAVYFCTRRGYDWYFDVWGQGTTVTV (SEQ ID NO: 1); and (ii) a variable light chain (VL) sequence with at least 90% sequence identity to DIQMTQSPSSLSASVGDRVITICRASQDINSYLSWFQQKPGKAPKTLIYRANRLESQVPSRFSGSG SGTDYTLTISSLQYEDFGIYYCQQYDESPWTFGGGTKLEIK (SEQ ID NO: 2). In some embodiments, the CFP further comprises an intracellular domain, wherein the intracellular domain comprises one or more intracellular

signaling domains, and wherein a wild-type protein comprising the intracellular domain does not comprise the extracellular domain. In some embodiments, the one or more intracellular signaling domains comprises a phagocytic signaling domain. In some embodiments, the phagocytosis signaling domain comprises an intracellular signaling domain derived from a receptor other than Megf10, MerTk, FcαR, and Bai1. In some embodiments, the phagocytosis signaling domain comprises an intracellular signaling domain derived from FcγR, FcαR or FcεR. In some embodiments, the phagocytosis signaling domain comprises an intracellular signaling domain with at least 90% sequence identity to LYCRRLKIQVRKAAITSYEKSDGVYTGLSTRNQETYETLKHEKPP (SEQ ID NO: 67). In some embodiments, the one or more intracellular signaling domains further comprises a proinflammatory signaling domain. In some embodiments, the proinflammatory signaling domain comprises a PI3-kinase (PI3K) recruitment domain. In some embodiments, the proinflammatory signaling domain comprises a sequence with at least 90% sequence identity to YEDMRGILYAAPQLRSIRGQPGPNHEEDADSY ENM (SEQ ID NO: 4). In some embodiments, the proinflammatory signaling domain is derived from an intracellular signaling domain of CD40. In some embodiments, the proinflammatory signaling domain comprises a sequence with at least 90% sequence identity to KVAKKPTNKAPHPKQEPQEINFPDDLPGSNTAAPVQETLHGCQPVTQEDGKESRISVQERQ (SEQ ID NO: 68). In some embodiments, the transmembrane domain comprises a CD8 transmembrane domain. In some embodiments, the transmembrane domain comprises a sequence with at least 90% sequence identity to IYIWAPLAGTCGVLLSLVIT (SEQ ID NO: 6). In some embodiments, the extracellular domain further comprises a hinge domain derived from CD8, wherein the hinge domain is operatively linked to the transmembrane domain and the CD5 binding domain. In some embodiments, the extracellular domain comprises a sequence with at least 90% sequence identity to ALSNSIMYFSHFVPVFLPAKPTTTPAPRPPTPAPTIASQPLSLRPEACRPAAGGAVHTRGLD (SEQ ID NO: 7). In some embodiments, the CFP comprises an extracellular domain comprising a scFv that specifically binds CD5, and a hinge domain derived from CD8; a hinge domain derived from CD28 or at least a portion of an extracellular domain from CD68; a CD8 transmembrane domain, a CD28 transmembrane domain or a CD68 transmembrane domain; and an intracellular domain comprising at least two intracellular signaling domains, wherein the at least two intracellular signaling domains comprise: a first intracellular signaling domain derived from FcγR or FcεR, and a second intracellular signaling domain comprising a PI3K recruitment domain, or derived from CD40. In some embodiments, the recombinant polynucleic acid is an mRNA or circRNA. In some embodiments, the nucleic acid is delivered into a myeloid cell. In some embodiments, the nucleic acid is delivered into a CD14⁺ cell, a CD14⁺CD16⁻ cell, an M0 macrophage, an M2 macrophage, an M1 macrophage or a mosaic myeloid cell/macrophage. In some embodiments, the fusion protein comprises a sequence with at least 90% sequence identity to

EIQLVQSGGGLVKPGGSVRISCAASGYTFTNYGMNWVRQAPGKGLEWMGWINTHTGEPTYAD
SFKGRFTFSLDDSKNTAYLQINSLRAEDTAVYFCTRRGYDWYFDVWGQGTTVTVSSGGGGSGG
GGSGGGGSDIQMTQSPSSLSASVGDRVTITCRASQDINSYLSWFQQKPGKAPKTLIYRANRLESG
VPSRFSGSGSGTDYTLTISSLQYEDFGIYYCQYDESPWTFGGGKLEIKSGGGGGSGALSNSIMYF

SHFVPVFLPAKPTTTPAPRPPTPAPTIASQPLSLRPEACRPAAGGAVHTRGLDIYIWAPLAGTCGV
 LLLSLVITLYCRRLKIQVRKAAITSYEKSDGVYTGLSTRNQETYETLKHEKPPQGSGSYEDMRGI
 LYAAPQLRSIRGQPGPNHEEDADSYENM (SEQ ID NO: 69). In some embodiments, the fusion
 protein comprises a sequence with at least 90% sequence identity to
 EIQLVQSGGGLVKPGGSVRISCAASGYTFTNYGMNWVRQAPGKGLEWMGWINTHTGEPTYAD
 SFKGRFTFSLDDSKNTAYLQINSLRAEDTAVYFCTRRGYDWYFDVWGQGTTVTVSSGGGGSGG
 GGSGGGGSDIQMTQSPSSLSASVGDRVTITCRASQDINSYLSWFQQKPGKAPKTLIYRANRLESG
 VPSRFSGSGSGTDYTLTISSLQYEDFGIYYCQQYDESPWTFGGGKLEIKSGGGGSGALSNSIMYF
 SHFVPVFLPAKPTTTPAPRPPTPAPTIASQPLSLRPEACRPAAGGAVHTRGLDIYIWAPLAGTCGV
 LLLSLVITLYCRRLKIQVRKAAITSYEKSDGVYTGLSTRNQETYETLKHEKPPQKKVAKKPTNKAP
 HPKQEPQEINFPDDLPGSNTAAPVQETLHGCQPVTQEDGKESRISVQERQ (SEQ ID NO: 70) or
 EIQLVQSGGGLVKPGGSVRISCAASGYTFTNYGMNWVRQAPGKGLEWMGWINTHTGEPTYAD
 SFKGRFTFSLDDSKNTAYLQINSLRAEDTAVYFCTRRGYDWYFDVWGQGTTVTVSSGGGGSGG
 GGSGGGGSDIQMTQSPSSLSASVGDRVTITCRASQDINSYLSWFQQKPGKAPKTLIYRANRLESG
 VPSRFSGSGSGTDYTLTISSLQYEDFGIYYCQQYDESPWTFGGGKLEIKSGGGGSGALSNSIMYF
 SHFVPVFLPAKPTTTPAPRPPTPAPTIASQPLSLRPEACRPAAGGAVHTRGLDIYIWAPLAGTCGV
 LLLSLVITLYCRRLKIQVRKAAITSYEKSDGVYTGLSTRNQETYETLKHEKPPQKKVAKKPTNKAP
 PHPKQEPQEINFPDDLPGSNTAAPVQETLHGCQPVTQEDGKESRISVQERQ (SEQ ID NO: 71).

[00453] In some embodiments, the fusion protein is a transmembrane protein, an intracellular protein or an intracellular protein. In one embodiment the fusion protein is directed to enhancing the function of an immune cell, e.g., a myeloid cell, selected from monocyte, macrophages dendritic cells or precursors thereof. In one embodiment the fusion protein augments a cellular function of an immune cell, such as phagocytosis. The disclosure is not limited by the transgenes that can be expressed using the methods and compositions described. The transgenes indicated in this section are exemplary.

[00454] Provided herein are exemplary transgene candidates, for stable integration into the genome of a phagocytic cell. In one embodiment the transgene is a recombinant nucleic acid encoding a phagocytic receptor (PR) fusion protein (CFP). The recombinant nucleic acid has a PR subunit comprising: (i) a transmembrane domain, and (ii) an intracellular domain comprising a phagocytic receptor intracellular signaling domain; and an extracellular antigen binding domain specific to an antigen of a target cell; wherein the transmembrane domain and the extracellular antigen binding domain are operatively linked such that antigen binding to the target by the extracellular antigen binding domain of the fused receptor activated in the intracellular signaling domain of the phagocytic receptor. In some embodiments, the recombinant nucleic acid encodes a chimeric antigen receptor. In some embodiments, the chimeric antigen receptor is a chimeric antigen receptor (phagocytosis) (CAR-P). In some embodiments, the fusion protein is a recombinant protein for locking anti-phagocytic signals. In some embodiments, the fusion protein is a phagocytosis enhancing chimeric protein. In some embodiments, the chimeric protein has intracellular domains comprising active phagocytosis signal transduction domains. In some embodiments, the chimeric protein enhances the phagocytic potential by enhancing the inflammatory potential of the phagocytic cell

ICD of human MARCO having an amino acid sequence:

MRNKKILKEDELLSETQQAAFHQIAMEPFENVPKPKRRNGVNF (SEQ ID NO: 72). In some embodiments, the PSR subunit comprises a variant which is at least 70%, 75%, 80%, 85%, 90% or 95% identical to the intracellular domain of MARCO.

[00462] In some embodiments, for example, the PSR (phagocytic scavenger receptor) comprises a transmembrane region of human MARCO.

[00463] In some embodiments, the recombinant nucleic acid encodes an intracellular domain of human SRA1. The PSR subunit comprises an intracellular domain having a 50 amino acid ICD of human SRA1 having an amino acid sequence: MEQWDHFHNQQEDTSCSESVKFDARSMTA

LLPPNPKNPSLQEKLKSFK (SEQ ID NO: 73). In some embodiments, the PSR subunit comprises a variant which is at least 70%, 75%, 80%, 85%, 90% or 95% identical to the intracellular domain of human SRA1. The intracellular region of SRA has a phosphorylation site.

[00464] In some embodiments, the PSR comprises a transmembrane region of human SRA1.

[00465] In some embodiments, for example, the recombinant nucleic acid comprises an intracellular domain of CD36. In some embodiments, the recombinant nucleic acid comprises a TM domain of CD36. Naturally occurring full length CD36 has two TM domains and two short intracellular domains, and an extracellular domain of CD36 binds to oxidized LDL. Both of the intracellular domains contain pairs of cysteines that are fatty acid acylated. It lacks known signaling domains (e.g. kinase, phosphatase, g-protein binding, or scaffolding domains). N-terminal cytoplasmic domain is extremely short (5–7 amino acid residues) and is closely associated with the internal leaflet of the plasma membrane. The carboxy-terminal domain contains 13 amino acids, containing a CXCX5K motif homologous to a region in the intracellular domain of CD4 and CD8 that is known to interact with signaling molecules. The intracellular domain of CD36 is capable of assembling a signaling complex that activates lyn kinases, MAP kinases and Focal Adhesion Kinases (FAK), and inactivation of src homology 2-containing phosphotyrosine phosphatase (SHP-2). Members of the guanine nucleotide exchange factors (GEFs) have been identified as potential key signaling intermediates.

[00466] In some embodiments, the recombinant nucleic acid encodes for example, an intracellular domain of human SCARA3. In some embodiments, the PSR subunit comprises a variant which is at least 70%, 75%, 80%, 85%, 90% or 95% identical to the intracellular domain of human SCARA3. In some embodiments, the PSR comprises the TM domain of SCARA3. In some embodiments, the TM domains are about 20-30 amino acids long.

[00467] Scavenger receptors may occur as homo or hetero dimers. MARCO, for example occurs as a homo trimer.

[00468] In some embodiments, the TM domain or the ICD domain of the PSP is not derived from FcR, Megf10, Bai1 or MerTK. In some embodiments, the ICD of the PSR does not comprise a CD3 zeta intracellular domain.

[00469] In some embodiments, the intracellular domain and transmembrane domains are derived from FcR beta.

PR at the extracellular domain. In some embodiments, the binding domain (e.g., scFv) and the extracellular domain of the PR are linked via a linker.

[00533] In some embodiments, the ECD antigen binding domain can bind to an intracellular antigen. In some embodiments, the intracellular antigen is a cancer antigen.

[00534] In some embodiments, the extracellular antigen binding domain binds to the target ligand with an affinity of less than 1000 nM. In some embodiments, the extracellular antigen binding domain binds to the target ligand with an affinity of less than 500 nM. In some embodiments, the extracellular antigen binding domain binds to the target ligand with an affinity of less than 450 nM. In some embodiments, the extracellular antigen binding domain binds to the target ligand with an affinity of less than 400 nM. In some embodiments, the extracellular antigen binding domain binds to the target ligand with an affinity of less than 350 nM. In some embodiments, the extracellular antigen binding domain binds to the target ligand with an affinity of less than 250 nM. In some embodiments, the extracellular antigen binding domain binds to the target ligand with an affinity of less than 200 nM. In some embodiments, the extracellular antigen binding domain binds to the target ligand with an affinity of less than 100 nM. In some embodiments, the extracellular antigen binding domain binds to the target ligand with an affinity ranging between than 200 nM to 1000 nM. In some embodiments, the extracellular antigen binding domain binds to the target ligand with an affinity ranging between than 300 nM to 1.5 mM. In some embodiments, the antigen binding domain binds to the target ligand with an affinity > 200 nM, > 300 nM or >500 nM.

Peptide Linker

[00535] In some embodiments, the extracellular antigen binding domains, scFvs are linked to the TM domain or other extracellular domains by a linker. In some embodiments, where there are more than one scfv at the extracellular antigen binding domain the more than scfvs are linked with each other by linkers.

[00536] In some embodiments, the linkers are flexible. In some embodiments, the linkers comprise a hinge region. Linkers are usually short peptide sequences. In some embodiments, the linkers are stretches of Glycine and one or more Serine residues. Other amino acids preferred for short peptide linkers include but are not limited to threonine (Thr), serine (Ser), proline (Pro), glycine (Gly), aspartic acid (Asp), lysine (Lys), glutamine (Gln), asparagine (Asn), and alanine (Ala) arginine (Arg), phenylalanine (Phe), glutamic acid (Glu). Of these Pro, Thr, and Gln are frequently used amino acids for natural linkers. Pro is a unique amino acid with a cyclic side chain which causes a very restricted conformation. Pro-rich sequences are used as interdomain linkers, including the linker between the lipoyl and E3 binding domain in pyruvate dehydrogenase (GA₂PA₃PAKQEA₃PAPA₂KAEAPA₃PA₂KA (SEQ ID NO: 75)). For the purpose of the disclosure, the empirical linkers may be flexible linkers, rigid linkers, and cleavable linkers. Sequences such as (G4S)_x (where x is multiple copies of the moiety, designated as 1, 2, 3, 4, and so on) (SEQ ID NO: 76) comprise a flexible linker sequence. Other flexible sequences used herein include several repeats of glycine, e.g., (Gly)₆ (SEQ ID NO: 77) or (Gly)₈ (SEQ ID NO: 78). On the other hand, a rigid linker may be used, for example, a linker (EAAAK)_x, where x is an integer, 1, 2, 3, 4 etc. (SEQ ID NO: 79) gives rise to a rigid linker.

[00537] In some embodiments, the linker comprises at least 2, or at least 3 amino acids. In some embodiments, the linker comprises 4 amino acids. In some embodiments, the linker comprises 5 amino

(graphically represented in **FIG. 4A and 4B**, and **FIG. 24**, construct **SEQ ID NO: 15**). The fused protein will specifically attach to the MS2-binding structural motif in the 3' UTR, and therefore any non-specific binding and retrotransposition is minimized (**FIG. 24**). GFP expression is monitored after transfection into HEK293T cells. Following a similar design, the ORF is fused with the protein binding sequences shown in left column of **Table 3** below, combined with a cognate sequence inserted in the 3'UTR region of the ORF2 shown in the corresponding right column in the same row.

Table 3 - Exemplary list of elements to enhance translation efficiency and stability of the LINE-1 proteins and increased expression of LINE-1 proteins.

Elements to be fused with the LINE-1 ORF2	3' UTR sequence recognizable by the element
PP7 coat protein	PP7
Streptavidin	S1m aptamer
Tobramycin	Tobramycin aptamer

[00681] *V. Modifying the endonuclease function of the retrotransposon.* In this example, the constructs are modified to test increase in GFP expression as follows. In a first experimental set, the LINE-1 plasmid GFP is cut at the 3' end of the endonuclease coding sequence of ORF2, and a sequence encoding the DNA binding domain (DBD) of a heterologous zinc finger protein (ZFP) is inserted. In another experimental set, the endonuclease domain is fused with a CRISPR nuclease. A variety of nucleases can be tested by modifying the LINE-1 plasmid GFP ORF by creating a fusion protein using DNA binding domains and cleavage domain as shown in a non-exhaustive list in **Table 4**. In addition, two ORF-2 domains are encoded in one set to facilitate dimerization. The construct that has higher GFP expression than the ORF2 endonuclease can be further selected. The plasmid designs are graphically represented in **FIG. 25**. GFP expression is monitored after transfection of the plasmids into HEK293T cells, and the set that yielded best.

Table 4 - Exemplary non-exhaustive list of additional DNA cleavage domains/enzymes that can be fused to or inserted in place of LINE-1 endonuclease.

Gene /Enzyme	Description
<i>FokI</i>	Class II endonuclease from <i>Flavobacterium okeanokoites</i> , recognition and cleavage sequence are separated by a few nucleotides; recognizes DNA sequence 5-GGATG-3'
Restriction enzymes,	e.g., HindII, EcoR1, BamH1
LAGLIDADG family nuclease A ("LAGLIDADG" disclosed as SEQ ID NO: 66)	Intron encoded homing proteins found in various genera including bacteria

GIY-YIG	This domain is found in the amino terminal region of exonuclease abc subunit c (uvrC), bacteriophage T4, endonuclease segA, segB, seg C, seg D, and seg E and group I introns of fungi and phage.
His-Cys box	Homing endonucleases containing two clusters of conserved histidine and cysteine residues over a 100 amino acid region.

the sequence encoding ORF1 of a LINE1-GFP construct (**FIG. 55A**), (iii) ORF2-N, in which an NLS was inserted at the N-terminal end of the sequence encoding ORF2 of a LINE1-GFP construct (**FIG. 56A**); (iv) ORF2-N, in which an SV40 NLS and a linker was inserted at the N-terminal end of the sequence encoding ORF2 of a LINE1-GFP construct (**FIG. 57A**); and (v) ORF2-C in which an NLS was inserted at the C-terminal end of the sequence encoding ORF2 of a LINE1-GFP construct (**FIG. 58A**). In each case, retrotransposition using the LINE GFP NLS constructs was measured in number of GFP insertions per cell (genome).

[00714] Table 12A details sequences relevant to the constructs in tests described for **FIGs 54A-55C**.

Table 12A

NLS	Amino acid sequence and nucleic acid sequence
(#3) SV40 NLS	PKKKRKV (SEQ ID NO: 80) ccaaagaagaagcggaagggtc (SEQ ID NO: 81)
(#4) SV40 NLS + Linker (linker sequence underlined)	PKKKRKV <u>GGGS</u> (SEQ ID NO: 82) ccaaagaagaagcggaagggtc <u>ggcgccggcgcagc</u> (SEQ ID NO: 83)
(#5) Nucleoplasmin NLS	KRPAATKKAGQAKKKK (SEQ ID NO: 84) aaaaggccggcgccacgaaaaaggccggccaggcaaaaaagaaaaag (SEQ ID NO: 85)
(#6) Nucleoplasmin NLS +Linker (linker sequence underlined)	KRPAATKKAGQAKKKK <u>GGGS</u> (SEQ ID NO: 86) aaaaggccggcgccacgaaaaaggccggccaggcaaaaaagaaaaag <u>ggcgccggcgcagc</u> (SEQ ID NO: 87)

[00715] As shown in **FIGs 54B-54C** and **FIGs 55B-55C**, NLS insertions at N-terminal and C-terminal of ORF1 sequence abolishes retrotransposition activity of the LINE-1 constructs. On the other hand, each of the constructs shown in **FIGs. 56A-58C** showed some improvement in retrotransposition, with insertion of the NLS in either ORF2 N or C terminal.

[00716] Table 12B details sequences relevant to the constructs in tests described for **FIGs 56-56C**.

Table 12B

NLS	Amino acid sequence and nucleic acid sequence
(#3) SV40 NLS	PKKKRKV (SEQ ID NO: 80) ccaaagaagaagcggaagggtc (SEQ ID NO: 81)
(#4) Nucleoplasmin NLS	KRPAATKKAGQAKKKK (SEQ ID NO: 84) aaaaggccggcgccacgaaaaaggccggccaggcaaaaaagaaaaag (SEQ ID NO: 85)
(#5) Linker + Nucleoplasmin NLS (linker sequence underlined)	<u>GGGSKRPAATKKAGQAKKKK</u> (SEQ ID NO: 88) <u>ggcgccggcgcagc</u> aaaaggccggcgccacgaaaaaggccggccaggcaaaaaagaaaaag (SEQ ID NO: 89)

[00717] With insertion of an NLS at the N terminal end of ORF2 sequence, about 2-fold increase in number of insertions per cell. (**FIGs. 56B, and 56C**). However, using the SV40 NLS with the addition of a linker sequence shown in **Table 12C**, increases the number of GFP integrations by about 3.5 folds (**FIGs. 57B and 57C**).

[00718] **Table 12C** details sequences relevant to the constructs in tests described for **FIGs 57-57C**.

Table 12C

NLS	Amino acid sequence and nucleic acid sequence
(#4a) SV40 NLS + Linker (linker sequence underlined)	PKKKRKV <u>VGGGS</u> (SEQ ID NO: 82) ccaaagaagaagcggaaggtc <u>ggcggcggcagc</u> (SEQ ID NO: 83)

[00719] However, surprisingly, insertion of the NLS at the ORF2 C terminus increases the retrotransposition by about 5 fold (**FIG. 58B-58C**).

[00720] **Table 12D** details sequences relevant to the constructs in tests described for **FIGs 58A-58C**.

Table 12D

NLS	Amino acid sequence and nucleic acid sequence
(#3) SV40 NLS	PKKKRKV (SEQ ID NO: 80) ccaaagaagaagcggaaggtc (SEQ ID NO: 81)
(#4) Linker + SV40 NLS	<u>GGGSPKKRKV</u> (SEQ ID NO: 90) <u>ggcggcggcagc</u> ccaaagaagaagcggaaggtc (SEQ ID NO: 91)
(#5) Nucleoplasmin NLS	KRPAATKKAGQAKKKK (SEQ ID NO: 84) aaaaggccggcgccacgaaaaaggccggccaggcaaaaaagaaaaag (SEQ ID NO: 85)
(#6) Linker + Nucleoplasmin NLS	<u>GGGSKRPAATKKAGQAKKKK</u> (SEQ ID NO: 88) <u>ggcggcggcagc</u> aaaaggccggcgccacgaaaaaggccggccaggcaaaaaagaaaaag (SEQ ID NO: 89)

[00721] Insertion of both the SV40 NLS and Nucleoplasmin NLS C-terminal to ORF2 leads to >5% GFP positive cells compared to less than 2% in cells that were electroporated with a LINE1-GFP mRNA construct without an NLS sequence. These results indicate that the ORF2 modifications with NLS increases retrotransposition efficiency. Most significantly, insertion at the C-terminal provides higher increase in retrotransposition of LINE1 elements.

[00722] Exemplary plasmids used for these experiments are shown in **Table 13**.

gccgtcgccgatgggggtgttctgctggtagtggtcgggcgagctgcacgctgccgtcctcgatgttg tggcggtatcttgaagttcaccttgatgcggttcttctgcttgctcggccatgatatagacgttggtggc tggtgtagttgtactccagcttggtgccccaggatgttgccgtcctccttgaagtcgatgcccttcag ctcgatgcggttcaccagggtgtgcgcccctcgaacttcacctcgggcggggtccttgtagttgccgtcg tccttgaagaagatggtgcgctcctggacgtagccttcgggcatggcggaacttgaagaagtcgtgct gcttcatgtggtcggggtagcgggtgaagcactgcacgccgtaggtcagggtggtcacgagggtggg ccagggcacgggcagcttgccgggtggtgcagatgaacttcagggtcagcttgccgttaggtggcatcg ccctcgccctcgccggacacgctgaacttggtggccggtttacgtcgccgtccagctcgaccaggatgg gcaccaccccggtgaacagctcctcgcccttgctcaccatggtggcgggatctgacggttcactaaa ccagctctgcttatatagacctcccacggtacacgcctaccgcccatttgcggtcaatgggdcggagt tgttacgacattttgaaaagtcgccgttgattttggtgccccaaacaaactccattgacgtcaatggg gtggagacttgaaaatccccgtgagtcaaacgcgtatccacgcccattgatgtactgccccaaacgcg atcaccatggtaatagcgatgactaatacgtagatgtactgccaagtaggaaaagtcaccataaggtca tgtactgggcataatgccaggcggggccatttaccgtcattgacgtcaatagggggcgtagttggcat atgatacacttgatgtactgccaagtgggcagtttaccgtaaatactccaccattgacgtcaatgg aaagtccttattggcggttactatgggaacatacgtcattattgacgtcaatgggdcgggggtcggttg gcggtcagccaggcggggccatttaccgtaagttatgtaacgGGCTGCTGCCGGCTCTGCGGCCCTCT TCCGCGTCTTCGCCTTCGCCCTCAGACGAGTCGGATCTCCCTTTGGGCCCGCTCCCCGCCTGTCTAG CTTGACTGACTGAGATACAGCGTACCTTCAGCTCACAGACATGATAAGATACATTGATGAGTTTGA CAAACCACAAC TAGAATGCAGTGAAAAAATGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTAT TTGTAACCATTATAAGCTGCAATAAACAAGTTAACAACAACAATTGCATTCATTTTATGTTTCAGGT TCAGGGGGAGGTGTGGGAGGTTTTTTTAAAGCAAGTAAACCTCTACAAATGTGGTATTGGCCCATCT CTATCGGTATCGTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGTGCCCTC GGGCCGATATGCTATCTACCGGCATTGGCGCAGAAAAAATGCCTGATGCGACGCTGCGCGTCTTAT ACTCCCACATATGCCAGATTACAGCAACGGATACGGCTTCCCCAACTTGCCCACTTCCATACGTGTCC TCCTTACCAGAAATTTATCCTTAAGGTCGTACAGTATCCTGCAGGCGATCTCGAATTCGATCAAG ACATTCCTTTAATGGTCTTTTTCTGGACACCACTAGGGGTGAGAAGTAGTTTCATCAAATTTCTTCCC TCCCTAATCTCATTGGTTACCTTGGGCTATCGAACTTAATTAAGCGATCTGCATCTCAATTAGTCA GCAACCATAGTCCCGCCCCTAAC TCCGCCCATCCCGCCCCCTAACTCCGCCCAGTTCCGCCCATCTC CGCCCCATCGCTGACTAATTTTTTTTTTATTTATGCAGAGGCCGAGGCCGCTCGGCTCTGAGCTATT CCAGAAGTAGTGAGGAGGCTTTTTTTGGAGGCCTAGGCTTTTGCAAAGGAGGTAGCCAACATGATTGA ACAAGATGGATTGCACGCAGGTTCTCCCGCCGCTTGGGTGGAGAGGCTATTGCGCTATGACTGGGCA CAACAGACAATCGGCTGCTCTGATGCCGCCGTGTTCCGGCTGTCAGCGCAGGGGCGCCCGTTCTTT TTGTCAAGACCGACCTGTCCGGTGCCCTGAATGAACTCCAGGACGAGGCAGCGCGCTATCGTGGCT GGCCACGACGGGCGTTCTTGGCGAGCTGTGCTCGACGTTGTCACTGAAGCGGGAAGGGACTGGCTG CTATTGGGCGAAGTGCCGGGGCAGGATCTCCTGTCTCATCTCACCTTGCTCCTGCCGAGAAAGTATCCA TCATGGCTGATGCAATGCGGCGGCTGCATACGCTTGATCCGGCTACCTGCCCATTCGACCACCAAGC GAAACATCGCATCGAGCGAGCACGTACTCGGATGGAAGCCGGTCTTGTCGATCAGGATGATCTGGAC GAAGAGCATCAGGGGCTCGCGCCAGCCGAAGTTCGCCAGGCTCAAGGCGCGGATGCCCGACGGCG AGGATCTCGTCGTGACCCACGGCGATGCCTGCTTGCCGAATATCATGGTGGAAAATGGCCGCTTTTC TGGATTTCATCGACTGTGGCCGGCTGGGTGTGGCGGACCGCTATCAGGACATAGCGTTGGCTACCCGT GATATTGCTGAAGAGCTTGGCGGCGAATGGGCTGACCGCTTCTCTAGTATGTAAGCCCTGTGCC CCGATTTCGACGCGCATCGCCTTCTATCGCCTTCTTGACGAGTTCTTCTAGTATGTAAGCCCTGTGCC TTCTAGTTGCCAGCCATCTGTTGTTTGGCCCTCCCCCGTGCCTTCTTGACCTTGAAGGTGCCACT CCCCTGTCTTTCTAATAAAATGAGGAAATTCATCGCATTGTCTGAGTAGGTGTCAATTCTATTC TGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGA TGCGGTGGGCTCTATGGTTAATTAACCAGTCAAGTCAGCTACTTGGCGAGATCGACTTGTCTGGGTT TCGACTACGCTCAGAATTGCGTCAGTCAAGTTCGATCTGGTCTTGTCTATTGCACCCGTTCTCCGAT TACGAGTTTTCATTTAAATCATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCG TTGCTGGCGTTTTTTCCATAGGCTCCGCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGA GGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCCTC TCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCTTTCTCCCTTCGGGAAGCGTGGCGCTT TCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTGCTTCGCTCCAAGCTGGGCTGTGTGC ACGAACCCCCCGTTTACGCCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGT AAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGC GGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCT GCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAACAAACCAC CGCTGGTAGCGGTGGTTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAGGATCTCAAGAA GATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGG TCATGAGATTATCAAAAAGGATCTTCACCTAGATCTTTTAAATTAATAAAGTGAAGTTTAAATCAAT CTAAAGTATATAGTAAGTAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCA GCGATCTGTCTATTTTCGTTTCATCCATAGTTGCATTTAAATTTCCGAAGTCTCCAAGGCCCTCGTCCG AAAATCTTCAAAC (SEQ ID NO: 92)
--

	ccggtgggtgcagatgaacttcagggtcagcttgccgtaggtggcatcgccctcgccctcgccggaca cgctgaacttgtggccggtttacgtcgccgtccagctcgaccaggatgggaccaccccggtgaacag ctcctcgcccttgctcaccatggtggcgggatctgacggttcactaaaccagctctgcttatataga cctcccaccgtacacgcctaccgcccatttgcgtcaatggggcgagggtgttacgacattttggaaa gtcccgttgattttggtgccccaaaaaactcccattgacgtcaatggggtggagacttggaaatccc cgtgagtcaaaaccgctatccacgcccattgatgtactgccccaaaccgcatcaccatggtaatagcga tgactaatacgtagatgtactgccaagtaggaaagtcccataaggtcatgtactgggcataatgcc ggcggggccatttaccgctcattgacgtcaatagggggcgctacttggcatatgatacacttgatgtact gccaagtgggcagtttaccgtaaaatactccaccattgacgtcaatggaaagtccctattggcggtta ctatgggaacatacgtcatttattgacgtcaatggggcgggggtcggttggcggtcagccagcgggcc atctaccgtaagttatgtaacgGGCCTGCTGCCGGCTCTGCCGGCTCTTCCGCGTCTTCGCCTTCGC CCTCAGACGAGTCGGATCTCCCTTTGGGCCGCTCCCCGCTGTCTAGCTTGACTGACTGAGATACA GCGTACCTTCAGCTCACAGACATGATAAGATACATTGATGAGTTTGGACAAACCACAAGTAAAGCTG AGTGAAGAAAAATGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAACCATATAAGCTG CAATAACAAGTTAACAACAACAATTGCATTCAATTTATGTTTCAGGTTTCAAGGGGAGGTGTGGGAG GTTTTTTTAAAGCAAGTAAACCTCTACAAATGTGGTATTGGCCCATCTCTATCGGTATCGTAGCATA ACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGTGCCCTCGGGCCGATTGCTATCTAC CGGCATTGGCGCAGAAAAAATGCCTGATGCGACGCTGCGCGTCTTATACTCCACATATGCCAGAT TCAGCAACGGATACGGCTTCCCCAAGTTGCCCACTTCCATACGTGTCTCTTACCAGAAATTTATC CTTAAGGTGCTCAGCTATCCTGCAGGCGATCTCTCGATTTCGATCAAGACATTCTTTAATGGTCTT TTCTGGACACCACTAGGGGTGAGAAGTAGTTTCATCAAACTTTCTTCCCTCCCTAATCTCATTGGTTA CCTTGGGCTATCGAACTTAATTAAGCGATCTGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCC TAACTCCGCCCATCCCGCCCCCTAACTCCGCCCAGTTCCGCCCATTCTCCGCCCCATCGCTGACTAAT TTTTTTTATTTATGCAGAGGCCGAGGCCGCTCGGCCCTCTGAGCTATTCCAGAAGTAGTGAGGAGGC TTTTTTGGAGGCTAGGCTTTTGCAGAGGAGGTAGCCAAACATGATTGAACAAGATGGATTGCACGCA GGTTCTCCCGCGCTTGGGTGGAGAGGCTATTCCGCTATGACTGGGCACACAGACAATCGGCTGCT CTGATGCCGCCGTGTTCCGGCTGTGACGCGAGGGGCGCCCGGTTCTTTTTGTCAAGACCGACCTGTC CGGTGCCCTGAATGAAGTCCAGGACGAGGCGAGCGCGGCTATCGTGGCTGGCCACGACGGGCGTTCT TGCGCAGCTGTGCTCGACGTTGTCACTGAAGCGGGAAGGGACTGGCTGCTATTGGGCGAAGTGCCGG GGCAGGATCTCCTGTCTCATCTCACCTTGCTCCTGCCGAGAAAGTATCCATCATGGCTGATGCAATGCG GCGGCTGCATACGCTTGATCCGGCTACCTGCCCATTCGACCACCAAGCGAAACATCGCATCGAGCGA GCACGTACTCGGATGGAAGCCGGTCTTGTGATCAGGATGATCTGGACGAAGAGCATCAGGGGCTCG CGCCAGCCGAAGTGTTCGCCAGGCTCAAGGCGCGGATGCCCGACGGCGAGGATCTCGTCTGACCCA CGGCGATGCCTGCTTGCCGAATATCATGGTGGAAAATGGCCGCTTTTCTGGATTATCGACTGTGGC CGGCTGGGTGTGGCGGACCGCTATCAGGACATAGCGTTGGCTACCCGTGATATTGCTGAAGAGCTTG GCGGCGAATGGGCTGACCGCTTCCCTCGTGCTTTACGGTATCGCCGCTCCCGATTTCGACGCGCATCGC CTTCTATCGCCTTCTTGACGAGTTCTTCTAGTATGTAAGCCCTGTGCCTTCTAGTTGCCAGCCATCT GTTGTTTGCCCCCTCCCCCGTGCTTCTTACCCCTGGAAGGTGCCACTCCCCTGTCCTTTCTTAAT AAAATGAGGAAATTCATCGCATTGTCTGAGTAGGTGTCAATCTATTCTGGGGGGTGGGGTGGGGCA GGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGTT AATTAACCAAGTCAAGTCAGCTACTTGGCGAGATCGACTTGTCTGGGTTTCGACTACGCTCAGAATTG CTGCAGTCAAGTTTCATCTGGTCTTGTCTATTGCACCCGTTCTCCGATTACGAGTTTCATTAAATC ATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCCGCTTGTCTGGCGTTTTCACATA GGCTCCGCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGG ACTATAAAGATAACAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCG CTTACCGGATACCTGTCCGCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTA GGTATCTCAGTTTCGGTGTAGGTGCTTCCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTACGCC CGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCA CTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGA AGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGT TACCTTCGGAAGAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTT TTTGTGTTGCAAGCAGCAGATTACGCGCAGAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTCTA CGGGGTCTGACGCTCAGTGGAACGAAAACCTACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAG GATCTTACCTAGATCCTTTTAAATTAATAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAA ACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTTCGTT CATCCATAGTTGCATTTAAATTTCCGAACCTCTCCAAGGCCCTCGTCGGAAAAATCTTCAAACCTTTTCG TCCGATCCATCTTGCAGGCTACCTCTCGAACGAACCTATCGCAAGTCTCTTGGCCGGCCTTGCGCCTT GGCTATTGCTTGGCAGCGCCTATCGCCAGGTATTACTCCAATCCCGAATATCCGAGATCGGGATCAC CCGAGCAAGTTCAACCTACATCCTCAATCCCGATCTATCCGAGATCCGAGGAATATCGAAATCGGG GCGCGCTGGTGTACCGAGAACGAT (SEQ ID NO: 93)
>RET-004 LINE1-	TCGAAATCGGGGCGCGCCTGGTGTACCGAGAACGATCCTCTCAGTGCGAGTCTCGACGATCCATATC GTTGCTTGGCAGTCAGCCAGTCGGAATCCAGCTTGGGACCCAGGAAGTCCAATCGTCAGATATTGTA CTCAAGCCTGGTCACGGCAGCGTACCGATCTGTTTTAAACCTAGATATTGATAGTCTGATCGGTCAAC GTATAATCGAGTCCTAGCTTTTGCAGAACATCTATCAAGAGACAGGATCAGCAGGAGGCTTTTCGCATG

	cctaccgcccatttgcgtcaatggggcggagttgttacgacattttggaaagtcccgttgattttggtgccccaaacaaactcccattgacgtcaatggggtggagacttggaaatccccgtgagtcacaaaccgctatccacgcccattgatgtactgccccaaaccgcatcaccatggtaatagcgatgactaatacgtagatgtactgccaagtaggaaagtcccataaggtcatgtactgggcataatgccaggcggggccatttaccgtcattgacgtcaatagggggcgtaacttggcatatgatacacttgatgtactgccaagtgggcagtttaccgtaataactccacccattgacgtcaatggaaagtccctattggcgttactatgggaacatacgtcattattgacgtcaatgggcgggggtcggttgggcgggtcagccaggcggggccatttaccgtaagttatgtaacgGGCCTGCTGCCGGCTCTGCGGCCTCTTCCGCGTCTTCGCCTTCGCCCTCAGACGAGTCGGA TCTCCCTTTGGGCCGCTCCCGCCTGTCTAGCTTAGCTGACTGAGATACAGCGTACCTTCAGCTCA CAGACATGATAAGATACATTGATGAGTTTGGACAAACCACAAGTAGAATGCAGTGAAAAAAATGCTT TATTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAAACCATTATAAGCTGCAATAAAACAAGTTAAC AACAACAATTGCATTTCATTTTATGTTTTAGGTTTCAAGGGGAGGTGTGGGAGGTTTTTTAAAGCAAGT AAAACCTCTACAAATGTGGTATTGGCCCATCTCTATCGGTATCGTAGCATAACCCCTTGGGGCCTCT AAACGGGTCTTGAGGGGTTTTTTGTGCCCTCGGGCCGGATTGCTATCTACCGGCATTGGCGCAGAA AAAAATGCCTGATGCGACGCTGCGCGTCTTATACTCCACATATGCCAGATTAGCAACGGATACGG CTTCCCCAAGTTGCCCACTTCCATACGTGTCTCTTACCAGAAATTTATCCTTAAGGTCGTCAGCT ATCCTGCAGGCGATCTCTCGATTTTCGATCAAGACATTCCTTTAATGGTCTTTTCTGGACACCACTAG GGGTCAGAAGTAGTTTCATCAAACCTTTCTTCCCTCCCTAATCTCATTGGTTACCTTGGGCTATCGAAA CTTAATTAAGCGATCTGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCCTAACTCCGCCCCATCCC GCCCCCTAACTCCGCCCCAGTTCCGCCCCATTCTCCGCCCCATCGCTGACTAATTTTTTTTATTTATGCA GAGGCCGAGGCCGCTCGGCCTCTGAGCTATTCCAGAAGTAGTGAGGAGGCTTTTTTGGAGGCCTAG GCTTTTGCAAAGGAGGTAGCCAACATGATTGAACAAGATGGATTGCACGCAGGTTCTCCCGCCGCTT GGGTGGAGAGGCTATTCCGGCTATGACTGGGCACAACAGACAATCGGCTGCTCTGATGCCGCCGTGTT CCGGCTGTGACGCGAGGGGCGCCCGGTTCTTTTTGTCAAGACCGACCTGTCCGGTGCCCTGAATGAA TCCAGGACGAGGCAGCGCGGCTATCGTGCTGGCCACGACGGGCGTTCCTTGCGCAGCTGTGCTCG ACCTTGTCTACTGAAGCGGGAAGGGAGCTGCTGCTATTGGGCGAAGTGCCGGGCGAGTATCCTGTCT ATCTCACCTTGCTCCTGCCGAGAAAGTATCCATCATAGGCTGATGCAATGCCGGCGGCTGCATACGCTT GATCCGGCTACCTGCCCATTCGACCACCAAGCGAAACATCGCATCGAGCGAGCACGTACTCGGATGG AAGCCGGTCTTGTCGATCAGGATGATCTGGACGAAGAGCATCAGGGGCTCGCGCCAGCCGAAGTGT CGCCAGGCTCAAGGCGCGGATGCCGACGCGGAGGATCTCGTCGTGACCCACGGCGATGCCTGCTTG CCGAATATCATGGTGGAAAATGGCCGCTTTTCTGGATTTCATCGACTGTGGCCGGCTGGGTGTGGCGG ACCGCTATCAGGACATAGCGTTGGCTACCCGTGATATTGCTGAAGAGCTTGGCGGCGAATGGGCTGA CCGCTTCCCTCGTGCTTTACGGTATCGCCGCTCCCGATTTCGACGCGCATCGCCTTCTATCGCCTTCTT GACGAGTTCTTCTAGTATGTAAGCCCTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCCTCCC CCGTGCTTCCCTTGACCCTGGAAGGTGCCACTCCCCTGTCTTCTTAATAAAATGAGGAAATTCG ATCGCATTTGCTGAGTAGGTGTCTATTCTTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAG GATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGTTAATTAACCAGTCAAGT CAGCTACTTGGCGAGATCGACTTGTCTGGGTTTTCGACTACGCTCAGAATTGCGTCAGTCAAGTTTCGA TCTGGTCTTGCTATTGCACCCGTTCTCCGATTACGAGTTTCATTTAAATCATGTGAGCAAAAAGGCC AGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCTGA CGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAG CCGCTTTCCCTTGGAAAGCTCCCTCGTGCGCTCTCTGTTCCGACCCCTGCCGCTTACCGGATACCTGT CCGCTTTCTCCCTTCGGAAGCGTGCGGCTTTCTCATAGCTACGCTGAGGTATCTCAGTTTCGGT GTAGGTCGTTTCGCTCCAAGCTGGGCTGTGTGACGAAACCCCCCGTTACGGCCGACCGCTGCGCCTTA TCCGGTAAGTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTG GTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAAGTGGTGGCCTAACTA CGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGA GTTGGTAGCTCTTGATCCGGCAAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTGTTGCAAGCAGC AGATTACGCGCAGAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCA GTGGAACGAAAACCTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATC CTTTTAAATTAATAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAACTTGGTCTGACAGTT ACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTTCATCCATAGTTGCATT TAAATTTCCGAAGTCTCCAAGGCCCTCGTCGGAAAAATCTTCAAACCTTTTCGTCCGATCCATCTTGCA GGCTACCTCTCGAACGAAGTATCGCAAGTCTCTTGGCCGGCCTTGCGCCTTGGCTATTGCTTGGCAG CGCCTATCGCCAGGTATTACTCCAATCCCGAATATCCGAGATCGGGATCACCCGAGAGAAAGTTCAAC CTACATCCTCAATCCCGATCTATCCGAGATCCGAGGAATA (SEQ ID NO: 94)
RET-005 LINE1-GFP ORF1-C Linker_Nucleop	TATCGAAATCGGGGCGCGCCTGGTGTACCGAGAACGATCCTCTCAGTGCGAGTCTCGACGATCCATA TCGTTGCTTGGCAGTCAGCCAGTCGGAATCCAGCTTGGGACCCAGGAAGTCCAATCGTCAGATATTG TACTCAAGCCTGGTCACGGCAGCGTACCGATCTGTTTAAACCTAGATATTGATAGTCTGATCGGTCA ACGTATAATCGAGTCTTAGCTTTTGCAAAACATCTCAAGAGACAGGATCAGCAGGAGGCTTTCGCA TGAGTATTCAACATTTCCGTGTCGCCCTTATTCCTTTTTTTCGCGCATTTTGCCCTTCTGTTTTGTC TCAACCCAGAAACGCTGGTGAAAGTAAAGATGCTGAAGATCAGTTGGGTGCGCGAGTGGGTACATC GAAGTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTTCGCCCCGAAGAACGCTTTCCAATGATGA GCATTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACGCCGGGCAAGAGCAACTCGG

	cggggccatttaccgctcattgacgtcaatagggggcgctacttggcatatgatacacttgatgtactgc caagtgggcagtttaccgtaaaatactccaccattgacgtcaatggaaagtcctattggcggttact atgggaacatacgtcattattgacgtcaatgggcgggggtcggtgggcgggtcagccaggcgggccat ttaccgtaagttatgtaacgGGCCTGCTGCCGGCTCTGCCGGCTCTTCCGCGTCTTCGCCTTCGCCC TCAGACGAGTCGGATCTCCCTTTGGGCCGCTCCCCGCTGTCTAGCTTGACTGACTGAGATACAGC GTACCTTCAGCTCACAGACATGATAAGATACATTGATGAGTTTGGACAAACCACAAC TAGAATGCAG TGAAAAAATGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAACCATTATAAGCTGCA ATAACAAGTTAACAACAACAATTGCATTCATTTTATGTTTCAGGTTTCAGGGGGAGGTGTGGGAGGT TTTTTAAAGCAAGTAAACCTCTACAAATGTGGATTTGGCCCATCTCTATCGGTATCGTAGCATAAC CCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGTGCCCTCGGGCCGATTGCTATCTACCG GCATTGGCGCAGAAAAAATGCCTGATGCGACGCTGCGCGTCTTATACTCCACATATGCCAGATTC AGCAACGGATACGGCTTCCCCAAGTTGCCACTTCCATACGTGTCTCTTACCAGAAATTTATCCT TAAGGTGCTGAGCTATCCTGCAGGCGATCTCTCGATTTTCGATCAAGACATTCTTTAATGGTCTTTT CTGGACACCACTAGGGGTCAGAAGTAGTTCATCAAACCTTTCTTCCCTCCCTAATCTCATTGGTTACC TTGGGCTATCGAACTTAATTAAGCGATCTGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCTA ACTCCGCCCATCCCGCCCCTAACTCCGCCAGTTCCGCCCATCTCCGCCCATCGCTGACTAATTT TTTTTATTTATGCAGAGGCCGAGGCCGCTCGGCCCTCTGAGCTATTCCAGAAGTAGTGAGGAGGCTT TTTTGGAGGCCTAGGCTTTTGCAAAGGAGGTAGCCAACATGATTGAACAAGATGGATTGCACGCAGG TTCTCCCGCCGCTTGGGTGGAGAGGCTATTTCGGCTATGACTGGGCACAACAGACAATCGGCTGCTCT GATGCCCGCGTGTTCGGCTGTGACGCGAGGGGCGCCCGGTCTTTTTGTCAAGACCGACCTGTCCG GTGCCCTGAATGAACTCCAGGACGAGGCGAGCGCGGCTATCGTGGCTGGCCACGACGGGCGTTCCCTTG CGCAGCTGTGCTCGACGTTGTCACTGAAGCGGGAAGGGACTGGCTGCTATTGGGCGAAGTGCCGGGG CAGGATCTCCTGTCTCATCTCACCTTGCTCCTGCCGAGAAAGTATCCATCATGGCTGATGCAATGCGGC GGCTGCATACGCTTGATCCGGCTACCTGCCCATTCGACCACCAAGCGAAACATCGCATCGAGCGAGC ACGTACTCGGATGGAAGCCGGTCTTGTCGATCAGGATGATCTGGACGAAGAGCATCAGGGGCTCGCG CCAGCCGAAGTGTTCGCCAGGCTCAAGGCGCGGATGCCCGACGGCGAGGATCTCGTGCAGCCACG GCGATGCCTGCTTGCCGAATATCATGGTGGAAAATGGCCGCTTTTCTGGATTTCATCGACTGTGGCCG GCTGGGTGTGGCGGACCGCTATCAGGACATAGCGTTGGCTACCCGTGATATTGCTGAAGAGCTTGGC GGCGAATGGGCTGACCGCTTCCCTCGTGCTTTACGGTATCGCCGCTCCCGATTTCGACGCGCATCGCCT TCTATCGCCTTCTTGACGAGTTCTTCTAGTATGTAAGCCCTGTGCCTTCTAGTTGCCAGCCATCTGT TGTTTTGCCCCCTCCCCCGTGCCTTCCCTTGACCCTGGAAGGTGCCACTCCCCTGTCTTTCCCTAATAA AATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCTATTCTATTCTGGGGGTGGGGTGGGGCAGG ACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGTTAA TTAACCAGTCAAGTCAGCTACTTGGCGAGATCGACTTGTCTGGGTTTCGACTACGCTCAGAATTGCG TCAGTCAAGTTCGATCTGGTCCTTGCTATTGCACCCGTTCTCCGATTACGAGTTTCATTTAAATCAT GTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGG CTCCGCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAAACCCGACAGGAC TATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCT TACCGGATACCTGTCCGCTTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGG TATCTCAGTTCGGTGTAGGTGCTTCCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTTCAGCCCG ACCGCTGCGCCTTATCCGGTAACCTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACT GGCAGCACCCTAGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAG TGGTGGCCTAAGTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTA CCTTCGGAAGAAAGAGTTGGTAGCTCTTGATCCGGCAAAACAAACCACCGCTGGTAGCGGTGTTTTT TGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACG GGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAAGGA TCTTCACCTAGATCCTTTTAAATTAATAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAAC TTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCTGTTCA TCCATAGTTGCATTTAAATTTCCGAACCTCTCCAAGGCCCTCGTCGAAAAATCTTCAAACCTTTCTGTC CGATCCATCTTGCAGGCTACCTCTCGAACGAACCTATCGCAAGTCTCTTGGCCGGCCTTGCGCCTTGG CTATTGCTTGGCAGCGCCTATCGCCAGGTATTACTCCAATCCCGAATATCCGAGATCGGGATCACCC GAGAGAAGTTCAACCTACATCCTCAATCCCGATCTATCCGAGATCCGAGGAA (SEQ ID NO: 95)
RET-006 LINE1-GFP ORF1-N SV40 NLS	GTCCGAAAATCTTCAAACCTTTTCGTCCGATCCATCTTGCAGGCTACCTCTCGAACGAACATATCGCAA GTCTCTTGGCCGGCCTTTCGCGCTTGGCTATTGCTTGGCAGCGCCTATCGCCAGGTATTACTCCAATC CCGAATATCCGAGATCGGGATCACCCGAGAGAAGTTCAAACCTACATCCTCAATCCCGATCTATCCGA GATCCGAGGAATATCGAAATCGGGGCGCGCCTGGTGTACCGAGAACGATCCTCTCAGTGCGAGTCTC GACGATCCATATCGTTGCTTGGCAGTCAGCCAGTCGGAATCCAGCTTGGGACCCAGGAAGTCCAATC GTCAGATATTGTACTCAAGCCTGGTCACGGCAGCGTACCGATCTGTTTAAACCTAGATATTGATAGT CTGATCGGTCAACGTATAATCGAGTCTTAGCTTTTGCAAACATCTATCAAGAGACAGGATCAGCAGG AGGCTTTTCGATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCCTT CCTGTTTTTGTCTACCCAGAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCGCGAG TGGGTTACATCGAACTGGATCTCAACAGCGGTAAAGATCCTTGAGAGTTTTCGCCCCGAAAGAACGCTT TCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACGCCGGGCAA GAGCAACTCGGTGCGCGCATACACTATTCTCAGAATGACTTGGTTGAGTATTCACCAGTCAAGAAAA

	acccattgacgtcaatggaaagtccctattggcggttactatgggaacatacgtcattattgacgtca atgggcgggggtcggttgggcggtcagccaggcgggccatttaccgtaagttatgtaacgGGCCTGCT GCCGGCTCTGCGGCCTCTTCCGCGTCTTCGCCTTCGCCCTCAGACGAGTCGGATCTCCCTTTGGGCC GCCTCCCCGCCTGTCTAGCTTGACTGACTGAGATACAGCGTACCTTCAGCTCACAGACATGATAAGA TACATTGATGAGTTTGGACAAACCACAACCTAGAATGCAGTGAAAAAATGCTTTATTTGTGAAATTT GTGATGCTATTGCTTTATTTGTAACCATTATAAGCTGCAATAAACAAGTTAACAACAACAATTGCAT TCATTTTATGTTTTAGGTTTCAGGGGGAGGTGTGGGAGGTTTTTTTAAAGCAAGTAAACCTCTACAAA TGTGGTATTGGCCCATCTCTATCGGTATCGTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAG GGGTTTTTTGTGCCCCCTCGGGCCGATTGCTATGTACCGGCATTGGCGCAGAAAAAATGCCTGATG CGACGCTGCGCGTCTTATACTCCACATATGCCAGATTTCAGCAACGGATACGGCTTCCCCAAGCTTGC CCACTTCCATACGTGTCTCCTTACCAGAAATTTATCCTTAAGGTCGTACAGCTATCCTGCAGGCGAT CTCTCGATTTTCGATCAAGACATTCTTTAATGGTCTTTTCTGGACACCACTAGGGGTGAGAAAGTAGT TCATCAAACCTTTCTTCCCTCCCTAATCTCATTGGTTACCTTGGGCTATCGAAACTTAATTAAGCGAT CTGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCCTAACTCCGCCCATCCCGCCCCCTAACTCCGC CCAGTTCCGCCCATTTCTCCGCCCATCGCTGACTAATTTTTTTTATTTATGCAGAGGCCGAGGCCGC CTCGGCCTCTGAGCTATTCCAGAAGTAGTGAGGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAGGA GGTAGCCAACATGATTGAACAAGATGGATTGCACGCAGGTTCTCCCGCCGCTTGGGTGGAGAGGCTA TTCGGCTATGACTGGGCACAACAGACAATCGGCTGCTCTGATGCCGCCGTGTTCCGGCTGTACGCGC AGGGGCGCCCGGTTCTTTTTGTCAAGACCGACCTGTCCGGTGCCCTGAATGAACTCCAGGACGAGGC AGCGCGGCTATCGTGGCTGGCCACGACGGGCGTTCTTGGCGAGCTGTGCTCGACGTTGTCACTGAA GCGGGAAGGGACTGGCTGCTATTGGGCGAAGTGCCGGGGCAGGATCTCCTGTCTCATCTCACCTTGCTC CTGCCGAGAAAGTATCCATCATGGCTGATGCAATGCGGCGGCTGCATACGCTTGATCCGGCTACCTG CCCATTTCGACCACCAAGCGAAACATCGCATCGAGCGAGCACGTACTCGGATGGAAGCCGGTCTTGTC GATCAGGATGATCTGGACGAAGAGCATCAGGGGCTCGCGCCAGCCGAAGTTCGCCAGGCTCAAGG CGCGGATGCCCGACGGCGAGGATCTCGTCTGACCCACGGCGATGCCTGCTGCCGAATATCATGGT GGAAATGGCCGCTTTTCTGGATTTCATCGATGTGCGCCGCTGGGTGTGGCGACCGCTATCAGGAC ATAGCGTTGGCTACCCGTGATATTGCTGAAGAGCTTGGCGGCGAATGGGCTGAGCGCTTCTCTGTC TTTACGGTATCGCCGCTCCCGATTTCGCAGCGCATCGCCTTCTATCGCCTTCTTGACGAGTTCTTCTA GTATGTAAGCCCTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGGCCCTCCCCCGTGCCTTCCTTG ACCCTGGAAGGTGCCACTCCCCTGTCTTTCTAATAAAATGAGGAAATGCATCGCATTTGTCTGA GTAGGTGTCAATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAA TAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGTTAATTAACCAGTCAAGTCAGTACTTGCGGA GATCGACTTGTCTGGGTTTCGACTACGCTCAGAATTGCGTCAGTCAAGTTCGATCTGGTCTTGCTA TTGCACCCGTTCTCCGATTACGAGTTTCATTTAAATCATGTGAGCAAAAGGCCAGCAAAAGGCCAGG AACCCTAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCTGACGAGCATCACAAAA ATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGG AAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCT TCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTTCGCT CCAAGCTGGGCTGTGTGCACGAACCCCCCGTTACGCCCCGACCGCTGCGCCTTATCCGGTAACCTATCG TCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGC AGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAA GAACAGATATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGAAAAAGAGTTGGTAGCTCTTG ATCCGGCAAAACCAACCCGCTGGTAGCGGTGGTTTTTTTGTGTTGCAAGCAGAGATTACGCGCAGA AAAAAAGGATCTCAAGAAGATCCTTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGGAACGAAAACT CACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAATA ATGAAGTTTTTAAATCAATCTAAAGTATATATGAGTAACTTGGTCTGACAGTTACCAATGCTTAATC AGTGAGGCACCTATCTCAGCGATCTGTCTATTTCTGTTTCATCCATAGTTGCATTTAAATTTCCGAACT CTCCAAGGCCCTC (SEQ ID NO: 96)
RET-007 LINE1-GFP ORF1-N SV40 NLS Linker	GGAAAATCTTCAAACCTTTTCGTCCGATCCATCTTGCAAGGCTACCTCTCGAACGAACTATCGCAAGTC TCTTGCCCGCCTTGCGCCTTGGCTATTGCTTGGCAGCGCCTATCGCCAGGTATTACTCCAATCCCG AATATCCGAGATCGGGATCACCCGAGAGAAGTTCAACCTACATCCTCAATCCCGATCTATCCGAGAT CCGAGGAATATCGAAATCGGGGCGCGCCTGGTGTACCGAGAACGATCCTCTCAGTGCGAGTCTCGAC GATCCATATCGTTGCTTGGCAGTCAGCCAGTCGGAATCCAGCTTGGGACCCAGGAAGTCCAATCGTC AGATATTGTACTCAAGCCTGGTCACGGCAGCGTACCGATCTGTTTTAAACCTAGATATTGATAGTCTG ATCGGTCAACGTATAATCGAGTCCTAGCTTTTGCAAACATCTATCAAGAGACAGGATCAGCAGGAGG CTTTTCGATGAGTATTCAACATTTCCGTGTGCGCCTTATTCCCTTTTTTGGCGCATTTTGCCTTCCT GTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCGCGAGTGG GTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGCTTTCC AATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACGCCGGGCAAGAG CAACTCGGTGCGCGCATACACTATTCTCAGAATGACTTTGGTTGAGTATTACCAGTACAGAAAAAGC ATCTTACGGATGGCATGACAGTAAGAGAATTATGCACTGCTGCCATAACCATGAGTGATAACACTGC GGCCAACTTACTTCTGACAACGATTGGAGGACCGAAGGAGCTAACCCTTTTTTGCACAACATGGGG GATCATGTAACGCGCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTG ACACCACGATGCCTGTAGCAATGGCAACAACCTTGCGTAAACTATTAAGTGGCGAACTACTTACTCT

	ATGATAAGATACATTGATGAGTTTGGACAAACCACAACCTAGAAATGCAGTGAAAAAATGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAACCATTATAAGCTGCAATAAACAAAGTTAACAAACAA CAATTGCATTCATTTTATGTTTCAGGTTTCAGGGGGAGGTGTGGGAGGTTTTTTTAAAGCAAGTAAAAAC CTCTACAAATGTGGTATTGGCCCATCTCTATCGGTATCGTAGCATAAACCCCTTGGGGCCTCTAAACG GGTCTTGAGGGGTTTTTTGTGCCCCCTCGGGCCGGATTGCTATCTACCGGCATTGGCGCAGAAAAAAA TGCTTGATGCGACGCTGCGCGTCTTATACTCCCACATATGCCAGATTAGCAACGGATACGGCTTCC CCAACTTGCCCACTTCCATACGTGTCTCCTTACCAGAAATTTATCCTTAAGGTGCTCAGCTATCCT GCAGGCGATCTCTCGATTTTCGATCAAGACATTCCTTTAATGGTCTTTTCTGGACACCACTAGGGGTC AGAAGTAGTTCATCAAACCTTTCTTCCCTCCCTAATCTCATTGGTTACCTTGGGCTATCGAAACTTAA TTAAGCGATCTGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCCTAACTCCGCCCATCCCGCCCC TAACTCCGCCCAGTTCGCCCCATTCTCCGCCCATCGCTGACTAATTTTTTTTTTATTTATGCAGAGGC CGAGGCCGCCTCGGCCTCTGAGCTATTCCAGAAGTAGTGAGGAGGCTTTTTTGGAGGCCCTAGGCTTT TGCAAAGGAGGTAGCCAACATGATTGAACAAGATGGATTGCACGCAGGTCTCCCGCCGCTTGGGTG GAGAGGCTATTTCGGCTATGACTGGGCACAACAGACAATCGGCTGCTCTGATGCCGCCGTGTTCCGGC TGTCAGCGCAGGGGCGCCCGGTTCTTTTTGTCAAGACCGACCTGTCCGGTGCCCTGAATGAACTCCA GGACGAGGCAGCGCGGCTATCGTGGCTGGCCACGACGGGCGTTCCTTGCGCAGCTGTGCTCGACGTT GTCCTGAAGCGGGAAGGGACTGGCTGCTATTGGGCGAAGTGCCGGGGCAGGATCTCCTGTCTATCTC ACCTTGCTCCTGCCGAGAAAGTATCCATCATGGCTGATGCAATGCCGCGGCTGCATACGCTTGATCC GGCTACCTGCCCATTCGACCACCAAGCGAAACATCGCATCGAGCGAGCACGTACTCGGATGGAAGCC GGTCTTGTCGATCAGGATGATCTGGACGAAGAGCATCAGGGGCTCGCGCCAGCCGAACTGTTTCGCCA GGCTCAAGGCGCGGATGCCCGACGGCGAGGATCTCGTCGTGACCCACGGCGATGCCTGCTTGCCGAA TATCATGGTGGAAAATGGCCGCTTTTCTGGATTTCATCGACTGTGGCCGGCTGGGTGTGGCGGACCGC TATCAGGACATAGCGTTGGCTACCCGTGATATTGCTGAAGAGCTTGGCGGCGAATGGGCTGACCGCT TCCTCGTGCTTTACGGTATCGCCGCTCCCGATTTCGACGCGCATCGCCTTCTATCGCCTTCTTGACGA GTTCTTCTAGTATGTAAGCCCTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCCTCCCCCGTG CCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCTCTTCTTAATAAAATGAGGAAATTCATCTCGC ATTGTCTGAGTAGGTGTCAATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTG GGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGTTAATTAACCAGTCAAGTCAGCT ACTTGCGGAGATCGACTTGTCTGGGTTTCGACTACGCTCAGAATTGCGTCAGTCAAGTTCGATCTGG TCCTTGCTATTGCACCCGTTCTCCGATTACGAGTTTCATTTAAATCATGTGAGCAAAAAGGCCAGCAA AAGGCCAGGAACCGTAAAAAGGCCGCTTGTGGCGTTTTTCCATAGGCTCCGCCCCCTGACGAGC ATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACAGGCGTT TCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCCTGCCGCTTACCGGATACCTGTCCGCC TTTCTCCCTTCGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTTCGGTGTAGG TCGTTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCCGACCGCTGCGCCTTATCCGG TAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAAC AGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCT AACTAGAAAGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGG TAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTTGTTTGCAAGCAGCAGATT ACGCGCAGAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGA ACGAAAATCAGCTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTT AAATTAATAAGTGAAGTTTAAATCAATCTAAAGTATATGAGTAAACTTGGTCTGACAGTTACCAA TGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTTCGTTTCATCCATAGTTGCATTTAAAT TTCCGAACCTCTCAAGGCCCTCGTC (SEQ ID NO: 97)
RET-008 LINE1-GFP ORF1-N Nucleoplasmin NLS	GGAAAATCTTCAAACCTTTTCGTCCGATCCATCTTGCAGGCTACCTCTCGAACGAACTATCGCAAGTC TCTTGCGCGCCTTGCGCCTTGGCTATTGCTTGGCAGCGCCTATCGCCAGGTATTACTCCAATCCCCG AATATCCGAGATCGGGATCACCCGAGAGAAGTTCAACCTACATCCTCAATCCCGATCTATCCGAGAT CCGAGGAATATCGAAATCGGGGCGCGCCTGGTGTACCGAGAACGATCCTCTCAGTGCGAGTCTCGAC GATCCATATCGTTGCTTGGCAGTCAGCCAGTCGGAATCCAGCTTGGGACCCAGGAAGTCCAATCGTC AGATATTGTACTCAAGCCTGGTCACGGCAGCGTACCGATCTGTTTAAACCTAGATATTGATAGTCTG ATCGGTCAACGTATAATCGAGTCCTAGCTTTTGCAAACATCTATCAAGAGACAGGATCAGCAGGAGG CTTTCGATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCT GTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCGCGAGTGG GTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTTCGCCCCGAAGAACGCTTTCC AATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACGCCGGGCAAGAG CAACTCGGTGCGCGCATACACTATTCTCAGAATGACTTGGTTGAGTATTCACCAGTCACAGAAAAGC ATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGAGTGATAACACTGC GGCCAACTTACTTCTGACAACGATTGGAGGACCGAAGGAGCTAACCCTTTTTTGACAACATGGGG GATCATGTAACCTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTG ACACCTAGATGCCTGTAGCAATGGCAACAACCTTTCGCTAAACTATTAAGTGGCACTACTTACTCT AGCTTCCCGCAACAGTTGATAGACTGGATGGAGCGGATAAAGTTGCAGGACCACTCTGCGCTCG GCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCA TTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGC AACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAACCG

	TGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTGTGCCCTCGGGCCGGATTGCTATCTACCGGCAT TGGCGCAGAAAAAATGCCTGATGCGACGCTGCGCGTCTTATACTCCACATATGCCAGATTACAGCA ACGGATACGGCTTCCCCAACTTGCCCACTTCCATACGTGTCTCTTACCAGAAATTTATCCTTAAG GTCGTACAGCTATCCTGCAGGCGATCTCTCGATTTTCGATCAAGACATTCTTTAATGGTCTTTTCTGG ACACCACTAGGGGTCAGAAGTAGTTTCATCAAACCTTTCTTCCCTCCCTAATCTCATTGGTTACCTTGG GCTATCGAACTTAATTAAGCGATCTGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCTAACCTC CGCCCATCCCGCCCCTAACCTCCGCCCAGTTCCGCCCATTCTCCGCCCCATCGCTGACTAATTTTTTTT TATTTATGCAGAGGCCGAGGCCGCTCGGCCCTGAGCTATTCCAGAAGTAGTGAGGAGGCTTTTTT GGAGGCTTAGGCTTTTGCAAAGGAGGTAGCCAACATGATTGAACAAGATGGATTGCACGCAGGTTCT CCCGCCGCTTGGGTGGAGAGGCTATTTCGGCTATGACTGGGCACAACAGACAATCGGCTGCTCTGATG CCGCCGTGTTCCGGCTGTCAGCGCAGGGGCGCCCGGTTCTTTTTGTCAAGACCGACCTGTCCGGTGC CCTGAATGAACTCCAGGACGAGGCGCGGCTATCGTGGCTGGCCACGACGGGCGTTCTTTCGCGCA GCTGTGCTCGACGTTGTCACTGAAGCGGGAAGGGACTGGCTGCTATTGGGCGAAGTGCCGGGGCAGG ATCTCCTGTCTCATCTCACCTTGCTCCTGCCGAGAAAGTATCCATCATGGCTGATGCAATGCGGCGGCT GCATACGCTTGATCCGGCTACCTGCCCATTTCGACCACCAAGCGAAACATCGCATCGAGCGAGCACGT ACTCGGATGGAAGCCGGTCTTGTCGATCAGGATGATCTGGACGAAGAGCATCAGGGGCTCGCGCCAG CCGAAGTGTTCGCCAGGCTCAAGGCGCGGATGCCCCGACGGCGAGGATCTCGTCGTGACCCACGGCGA TGCCTGCTTGCCGAATATCATGGTGGAAAATGGCCGCTTTTCTGGATTTCATCGACTGTGGCCGGCTG GGTGTGGCGGACCGCTATCAGGACATAGCGTTGGCTACCCGTGATATTGCTGAAGAGCTTGGCGGCG AATGGGCTGACCGCTTCTCGTGCTTTACGGTATCGCCGCTCCCGATTTCGCAGCGCATCGCCTTCTA TCGCCTTCTTGACGAGTTCTTCTAGTATGTAAGCCCTGTGCCTTCTAGTTGCCAGCCATCTGTTGTT TGCCCCCTCCCCCGTGCTTCTTTCGACCTGGAAGGTGCCACTCCCACTGTCTTTCTAATAAAATG AGGAAATTGCATCGCATTGTCTGAGTAGGTGTCACTTCTATTCTGGGGGTGGGGTGGGGCAGGACAG CAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGTTAATTAA CCAGTCAAGTCAGCTACTTGGCGAGATCGACTGTCTGGGTTTCGACTACGTCAGAATTGCGTCAG TCAAGTTCGATCTGGTCTTGTCTATTGCACCCGTTCTCCGATTACGAGTTTCATTAAATCATGTTGA GCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCTTGCTGGCGTTTTTCCATAGGCTCC GCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAAACCCGACAGGACTATA AAGATACCAGGCGTTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACC GGATACCTGTCCGCTTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATC TCAGTTCCGGTGTAGGTGCTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTTCAGCCCGACCG CTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCA GCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGT GGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTT CGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTTGTT TGCAAGCAGCAGATTACGCGCAGAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGT CTGACGCTCAGTGGAACGAAAACCTACGTTAAGGGATTTTGGTTCATGAGATTATCAAAAAAGGATCTT CACCTAGATCCTTTTAAATTAATAAAGTAAAAATCAATCTAAAGTATATATGAGTAAACTTGG TCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTTCGTTTCATCCA TAGTTGCATTTAAATTTCCGAACCTCTCCAAGGCCCTCGTC (SEQ ID NO: 98)
RET-009 LINE1-GFP ORF1-N Nucleoplasmin NLS Linker	CAAACCTTTTCGTCCGATCCATCTTGCAGGCTACCTCTCGAACGAACTATCGCAAGTCTCTTGGCCGG CCTTGCGCCTTTGGCTATTGCTTGGCAGCGCCTATCGCCAGGTATTACTCCAATCCCGAATATCCGAG ATCGGGATCACCCGAGAGAAGTTCAACCTACATCCTCAATCCCGATCTATCCGAGATCCGAGGAATA TCGAAATCGGGGCGCGCCTGGTGTACCCGAGAACGATCCTCTCAGTGCAGTCTCGACGATCCATATC GTTGCTTGGCAGTCAGCCAGTCGGAATCCAGCTTGGGACCCAGGAAGTCCAATCGTCAGATATTGTA CTCAAGCCTGGTCACGGCAGCGTACCGATCTGTTTTAAACCTAGATATTGATAGTCTGATCGGTCAAC GTATAATCGAGTCCTAGCTTTTGCAAACATCTATCAAGAGACAGGATCAGCAGGAGGCTTTTCGCATG AGTATTCAACATTTCCGTGTGCGCCTTATTCCTTTTTTTCGGGCATTTTGCCTTCTGTTTTTGTCTC ACCCAGAAACGCTGGTGAAAGTAAAGATGCTGAAGATCAGTTGGGTGCGCGAGTGGGTTACATCGA ACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTTCGCCCCGAAGAACGCTTTCCAATGATGAGC ACTTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACGCCGGGCAAGAGCAACTCGGTC GCCGCATACACTATTCTCAGAATGACTTGGTTGAGTATTACACAGTCACAGAAAAGCATCTTACGGA TGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTA CTTCTGACAACGATTGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAA CTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGAT GCCTGTAGCAATGGCAACAACCTTGCGTAAACTATTAAGTGGCGAACTACTTACTCTAGCTTCCCCGG CAACAGTTGATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGG CTGGCTGGTTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTCAGCACT GGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGAT GAACGAAATAGACAGATCGCTGAGATAGGTGCTCAGTATTAAGCATTTGTAACCGATTCTAGGTG CATTTGGCGCAGAAAAAATGCCTGATGCGACGCTGCGCGTCTTATACTCCACATATGCCAGATTCA GCAACGGATACGGCTTCCCAACTTGCCCACTTCCATACGTGTCCTCCTTACCAGAAAATTTATCCTT AAGATCGTTTTAACTCGACTCTGGCTCTATCGAATCTCCGTGCTTTTCGAGCTTACGCGAACAGCCGT GGCGCTCATTTGCTCGTCGGGCATCGAATCTCGTCAGCTATCGTCAGCTTACCTTTTTTGGCAGCGAT

	GGACACCACTAGGGGTCAGAAGTAGTTCATCAAACCTTTCTTCCCTCCCTAATCTCATTGGTTACCTT GGGCTATCGAACTTAATTAAGCGATCTGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCTAAC TCCGCCCATCCCGCCCCCTAACTCCGCCCAGTTCCGCCCATTCTCCGCCCCATCGCTGACTAATTTTTT TTTATTTATGCAGAGGCCGAGGCCGCTCGGCCTCTGAGCTATTCCAGAAGTAGTGAGGAGGCTTTT TTGGAGGCCTAGGCTTTTGCAAAGGAGGTAGCCAACATGATTGAACAAGATGGATTGCACGCAGGTT CTCCCGCCGCTTGGGTGGAGAGGCTATTCCGCTATGACTGGGCACAACAGACAATCGGCTGCTCTGA TGCCGCCGTGTTCCGGCTGTCAGCGCAGGGGCGCCCGGTTCTTTTTGTCAAGACCGACCTGTCCGGT GCCCTGAATGAACCTCCAGGACGAGGCAGCGCGCTATCGTGGCTGGCCACGACGGGCGTTCCCTGCG CAGCTGTGCTCGACGTTGTCACCTTGCTCCTGCCGAGAAAGTATCCATCATGGCTGATGCAATGCCGCGG GGATCTCCTGTCTCATCTCACCTTGCTCCTGCCGAGAAAGTATCCATCATGGCTGATGCAATGCCGCGG CTGCATACGCTTGATCCGGCTACCTGCCCATTTCGACCACCAAGCGAAACATCGCATCGAGCGAGCAC GTACTCGGATGGAAGCCGGTCTTGTCGATCAGGATGATCTGGACGAAGAGCATCAGGGGCTCGCGCC AGCCGAACCTGTTTCGCCAGGCTCAAGGCGCGGATGCCCAGCGCGAGGATCTCGTCGTGACCCACGGC GATGCCTGCTTGCCGAATATCATGGTGGAAAATGGCCGCTTTTCTGGATTTCATCGACTGTGGCCGGC TGGGTGTGGCGGACCGCTATCAGGACATAGCGTTGGCTACCCGTGATATTGCTGAAGAGCTTGGCGG CGAATGGGCTGACCGCTTCCTCGTGCTTTACGGTATCGCCGCTCCCGATTTCGACGCGCATCGCCTTC TATCGCCTTCTTGACGAGTTCTTCTAGTATGTAAGCCCTGTGCCTTCTAGTTGCCAGCCATCTGTTG TTTGCCCTCCCGCTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCCTGTCTTTCTTAATAAAA TGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCAATTCTATTCTGGGGGTGGGGTGGGGCAGGAC AGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGTTAATT AACCAGTCAAGTCAGCTACTTGGCGAGATCGACTTGTCTGGGTTTCGACTACGCTCAGAATTGCGTC AGTCAAGTTCGATCTGGTCCTTGCTATTGCACCCGTTCTCCGATTACGAGTTTCATTTAAATCATGT GAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCT CCGCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTA TAAAGATACAGGCGTTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCCTGCCGCTTA CCGATACCTGTCCGCTTTCTCCCTTCGGGAAGCGCTGGCGCTTTCTCATCGCTACGCTGTAGGTA TCTCAGTTCCGTGTAGGTGCTTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCGTTTCAGCCGAC CGCTGCGCCTTATCCGGTAACCTATCGTCTTGAGTCCAACCCGTAAGACACGACTTATCGCCACTGG CAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAAGTG GTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACC TTCGGA AAAAGAGTTGGTAGCTCTTGATCCGGCAAAACAAACCCGCTGGTAGCGGTGGTTTTTTTG TTTGCAAGCAGCAGATTACGCGCAGAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGG GTCTGACGCTCAGTGAACGAAAACCTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATC TTCACCTAGATCCTTTTTAAATTA AAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAACTT GGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCTGTTTCATC CATAGTTGCATTTAAATTTCCGAACCTCTCCAAGGCCCTCGTCGGAAAAATCTT (SEQ ID NO: 99)
RET-010 LINE1-GFP ORF2-N SV40-NLS	GGGGCGCGCCTGGTGTACCGAGAACGATCCTCTCAGTGCGAGTCTCGACGATCCATATCGTTGCTTG GCAGTCAGCCAGTCGGAATCCAGCTTGGGACCCAGGAAGTCCAATCGTCAGATATTGTA CTCAAGCC TGGTCACGGCAGCGTACCGATCTGTTTAAACCTAGATATTGATAGTCTGATCGGTCAACGTATAATC GAGTCCTAGCTTTTGCAAACATCTATCAAGAGACAGGATCAGCAGGAGGCTTTCGCATGAGTATTCA ACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCTTCTGTTTTGCTCACCCAGAA ACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCGCGAGTGGGTACATCGAACTGGATC TCAACAGCGGTAAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGCTTTCCAATGATGAGCACTTTTAA AGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACGCCGAGCAAGCACTCGGTGCGCGCATATA CACTATTCTCAGAATGACTTGTTGAGTATTACAGTCACAGAAAAGCATCTTACGGATGGCATGA CAGTAAGAGAATTATGCAGTGCTGCCATAACCATGAGTGATAA CACTGCGGCCA ACTTACTTCTGAC AACGATTGGAGGACCGAAGGAGCTAACCGCTTTTTTGCAACATGGGGGATCATGTA ACTCGCCTT GATCGTTGGGAACCGGAGCTGAATGAAGCCATAACAAACGACGAGCGTGACACCAGATGCCTGTAG CAATGGCAACAACCTTGCGTAAACTATTA ACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAGTT GATAGACTGGATGGAGGCGGATAAAGTTGCAGGACC ACTTCTGCGCTCGGCCCTTCCGGCTGGCTGG TTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAG ATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAA TAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAACCGATTCTAGGTGCATTGGCG CAGAAAAAATGCCTGATGCGACGCTGCGCGTCTTATACTCCCACATATGCCAGATTTCAGCAACGGA TACGGCTTCCCCAACTTGCCCACTTCCATACGTGTCTCTCTTACCAGAAAATTTATCCTTAAGATCGT TTAAACTCGACTCTGGCTCTATCGAATCTCCGTCGTTTCGAGCTTACGCGAACAGCCGTGGCGCTCA TTTGCTCGTCGGGCATCGAATCTCGTCAGCTATCGTCAGCTTACCTTTTTTGGCAGCGATCGCGGCTC CCGACATCTTGGACCATTAGCTCCACAGGTATCTTCTTCCCTCTAGTGGTCATAACAGCAGCTTCAG CTACCTCTCAATTCAAAAAACCCCTCAAGACCCGTTTAGAGGGCCCAAGGGGTATGCTATCAATCG TTGCGTTACACACACAAAAAACCAACACACATCCATCTTCGATGGATAGCGATTTTATTATCTAACT GCTGATCGAGGTAGCCAGATCTAGTAATCAATTAACGGGGTCATTAGTTTCATAGCCCATATATGGAG TTCCGCGTTACATAACTTACGGTAAATGGCCCGCTGGCTGACCGCCCAACGACCCCGCCCATTTGA CGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGA GTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATT

	CACGACGGGCGTTCTTGCGCAGCTGTGCTCGACGTTGTCACTGAAGCGGGAAGGGACTGGCTGCTA TTGGGCGAAGTGCCGGGGCAGGATCTCCTGTCTATCTCACCTTGCTCCTGCCGAGAAAGTATCCATCA TGGCTGATGCAATGCGGGCGGCTGCATACGCTTGATCCGGCTACCTGCCCATTGACCACCAAGCGAA ACATCGCATCGAGCGAGCACGTA CTGGATGGAAGCCGGTCTTGTCGATCAGGATGATCTGGACGAA GAGCATCAGGGGCTCGCGCCAGCCGA ACTGTTGCCAGGCTCAAGGCGCGGATGCCCAGCGCGAGG ATCTCGTCGTGACCCACGGCGATGCCTGCTTGCCGAATATCATGGTGGA AATGGCCGCTTTTCTGG ATTCATCGACTGTGGCCGGCTGGGTGTGGCGGACCGCTATCAGGACATAGCGTTGGCTACCCGTGAT ATTGCTGAAGAGCTTGCGGGCGAATGGGCTGACCGCTTCCTCGTGCTTTACGGTATCGCCGCTCCCG ATTGCGAGCGCATCGCCTTCTATCGCCTTCTTGACGAGTTCTTCTAGTATGTAAGCCCTGTGCCTTC TAGTTGCCAGCCATCTGTTGTTTGCCCTCCCCCGTGCTTCCTTGACCCTGGAAGGTGCCACTCCC ACTGTCCTTTTCTAATAAAATGAGGAAATTCATCGCATTGTCTGAGTAGGTGTCAATTCTATTCTGG GGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGC GGTGGGCTCTATGGTTAATTAACCAGTCAAGTCAGCTACTTGCGGAGATCGACTTGTCTGGGTTTCG ACTACGCTCAGAATTGCGTCAGTCAAGTTCGATCTGGTCCTTGCTATTGCACCCGTTCTCCGATTAC GAGTTTCATTTAAATCATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTG CTGGCGTTTTTTCATAGGCTCCGCCCCCTGACGAGCATCACAAAATCGACGCTCAAGTCAGAGGT GGCGAAACCCGACAGGACTATAAAGATAACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCC TGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCTTTCTCCCTTCGGGAAGCGTGCGCTTTCT CATAGCTCACGCTGTAGGTATCTCAGTTCCGGTGTAGGTGCTTCGCTCCAAGCTGGGCTGTGTGCACG AACCCCCCGTTAGCCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAG ACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGT GCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCG CTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGC TGGTAGCGGTGGTTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAGGATCTCAAGAAGAT CCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTACGTTAAGGGATTTTGGTCA TGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTA AAAATGAAGTTTAAATCAATCTA AAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCATCTATCTCAGCG ATCTGTCTATTTTCGTTTCATCCATAGTTGCATTTAAATTTCCGA ACTCTCCAAGGCCCTCGTCGGAAA ATCTTCAAACCTTTTCGTCCGATCCATCTTGCAGGCTACCTCTCGAACGA ACTATCGCAAGTCTCTTG GCCGGCCTTGCGCCTTGCTATTGCTTGGCAGCGCCTATCGCCAGGTATTACTCCAATCCCGAATAT CCGAGATCGGGATCACCCGAGAGAAGTTCAACCTACATCCTCAATCCCGATCTATCCGAGATCCGAG GAATATCGAAATC (SEQ ID NO: 100)
ret-011- line1- gfp- orf2-n- sv40- nls- linker	TGGCCGGCCTTGCGCCTTGCTATTGCTTGGCAGCGCCTATCGCCAGGTATTACTCCAATCCCGAAT ATCCGAGATCGGGATCACCCGAGAGAAGTTCAACCTACATCCTCAATCCCGATCTATCCGAGATCCG AGGAATATCGAAATCGGGGCGCGCCTGGTGTACCGAGAACGATCCTCTCAGTGCGAGTCTCGACGAT CCATATCGTTGCTTGGCAGTCAGCCAGTCGGAATCCAGCTTGGGACCCAGGAAGTCCAATCGTCA TATTGTACTCAAGCCTGGTCACGGCAGCGTACCGATCTGTTTAAACCTAGATATTGATAGTCTGATC GGTCAACGTATAATCGAGTCCTAGCTTTTGCAAACATCTATCAAGAGACAGGATCAGCAGGAGGCTT TCGCATGAGTATTCAACATTTCCGTGTGCGCCTTATTCCCTTTTTTGCGGCATTTTGCTTCTCTGTT TTTGCTCACCCAGAAACGCTGGTGAAAGTAAAGATGCTGAAGATCAGTTGGGTGCGCGAGTGGGTT ACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTGCCCCGAAGAACGCTTTCCAAT GATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACGCCGGGCAAGAGCAA CTCGGTGCGGCATACACTATTCTCAGAATGACTTGGTTGAGTATTACCAAGTCACAGAAAAAGCATC TTACGGATGGCATGACAGTAAGAGAATTATGCACTGCTGCCATAACCATGAGTGATAACACTGCGGC CAACTTACTTCTGACAACGATTGGAGGACCGAAGGAGCTAACCGCTTTTTTGCAACATGCGGGAT CATGTAACCTGCGCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACA CCACGATGCCTGTAGCAATGGCAACAACCTTGCCTAACTATTAAGTGGCGAACTACTTACTCTAGC TTCCCGGCAACAGTTGATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCC CTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTG CAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAAC TATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAACCGATT CTAGGTGCATTGGCGCAGAAAAAATGCCTGATGCGACGCTGCGCGTCTTATACTCCACATATGCC AGATTGAGCAACGATACGGCTTCCCCA ACTTGCCCACTTCCATACGTGTCTCTTACCAGAAAT TATCCTTAAGATCGTTTAACTCGACTCTGGCTCTATCGAATCTCCGTGCTTTTCGAGCTTACGCGAA CAGCCGTGGCGCTCATTTGCTCGTGGGCATCGAATCTCGTCAGCTATCGTCAGCTTACCTTTTTTG CAGCGATCGCGGCTCCCGACATCTTGGACCATTAGCTCCACAGGTATCTTCTTCCCTCTAGTGGTCA TAACAGCAGCTTCAGCTACCTCTCAATTCAAAAAACCCCTCAAGACCCGTTTAGAGGGCCCCAAGGGG TTATGCTATCAATCGTTGCGTTACACACACAAAAACCAACACACATCCATCTTCGATGGATAGCGA TTTTATTATCTAACTGCTGATCGAGTGTAGCCAGATCTAGTAATCAATTACGGGGTCATTAGTTTCAT AGCCCCATATAGGAGTTCCGCGTTACATACTACGGTAAATGGCCCCGCTGGCTGACCGCCCAACG ACCCCCATGACATTGACGTCATAAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTG ACGTCAATGGGTGGAGTATTTACGGTAAACTGCCACTTGGCAGTACATCAAGTGTATCATATGCCA AGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCCGCTGGCATTATGCCAGTACATGACCT TATGGGACTTTCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGCTGATGCGGTT

	TTGCTCCTGCCGAGAAAGTATCCATCATGGCTGATGCAATGCGGCGGCTGCATACGCTTGATCCGGC TACCTGCCCATTTCGACCACCAAGCGAAACATCGCATCGAGCGAGCACGTACTCGGATGGAAGCCGGT CTTGTGCGATCAGGATGATCTGGACGAAGAGCATCAGGGGCTCGCGCCAGCCGAACTGTTCCGCCAGGC TCAAGGCGCGGATGCCCCGACGGCGAGGATCTCGTCGTGACCCACGGCGATGCCTGCTTGCCGAATAT CATGGTGGAAAATGGCCGCTTTTCTGGATTTCATCGACTGTGGCCGGCTGGGTGTGGCGGACCGCTAT CAGGACATAGCGTTGGCTACCCGTGATATTGCTGAAGAGCTTGCGGCGGAATGGGCTGACCGCTTCC TCGTGCTTTACGGTATCGCCGCTCCCGATTTCGACGCGCATCGCCTTCTATCGCCTTCTTGACGAGTT CTTCTAGTATGTAAGCCCTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCCTCCCCGTGCCT TCCTTGACCCTGGAAGGTGCCACTCCCACCTGTCTTTTCCCTAATAAAAATGAGGAAATGCATCGCAT GTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGA AGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGTTAATTAACCAGTCAAGTCAGCTACT TGGCGAGATCGACTTGTCTGGGTTTTCGACTACGCTCAGAATTGCGTCAGTCAAGTTCGATCTGGTCC TTGCTATTGCACCCGTTCTCCGATTACGAGTTTCATTTAAATCATGTGAGCAAAAAGGCCAGCAAAAAG GCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCTGACGAGCATC ACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCC CCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCTTT CTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTGCG TTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTTCAGCCCGACCGCTGCGCCTTATCCGGTAA CTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGG ATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACA CTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAG CTCTTGATCCGGCAAACAAACCCGCTGGTAGCGGTGGTTTTTTTTGTTTGCAAGCAGCAGATTACG CGCAGAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAAACG AAAACCTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAA TTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATAGTAAACTTGGTCTGACAGTTACCAATGC TTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTTCGTTTCATCCATAGTTGCATTTAAATTT CGAACTCTCCAAGGCCCTCGTCGAAAAATCTTCAAACCTTTTCGTCCGATCCATCTTGACGGCTACCT CTCGAACGAACCTATCGCAAGTCTCT (SEQ ID NO: 101)
RET-012 LINE1-GFP ORF2-N Nucleoplasmin-NLS	GACGATCCATATCGTTGCTTGGCAGTCAGCCAGTCGGAATCCAGCTTGGGACCCAGGAAGTCCAATC GTCAGATATTGTACTCAAGCCTGGTCACGGCAGCGTACCGATCTGTTTAAACCTAGATATTGATAGT CTGATCGGTCAACGTATAATCGAGTCCTAGCTTTTGCAAACATCTATCAAGAGACAGGATCAGCAGG AGGCTTTTCGATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTT CCTGTTTTTGTCTACCCAGAAACGCTGGTGAAAGTAAAGATGCTGAAGATCAGTTGGGTGCGCGAG TGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTGCCCCGAAGAACGCTT TCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACGCCGGGCAA GAGCAACTCGGTGCGCGCATACACTATTCTCAGAATGACTTGGTTGAGTATTCACCAGTCACAGAAA AGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGAGTGATAACAC TGCGGCCAACTTACTTCTGACAACGATTGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATG GGGGATCATGTAACCTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGC GTGACACCACGATGCCTGTAGCAATGGCAACAACCTTGCGTAAACTATTAACCTGGCGAACTACTTAC TCTAGCTTCCCGGCAACAGTTGATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGC TCGGCCCTTCCGGCTGGCTGGTTTTATTGCTGATAAACTCTGGAGCCGGTGAGCGTGGGTCTCGCGGTA TCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCA GGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAAGCATTTGGTAA CCGATTCTAGGTGCATTGGCGCAGAAAAAATGCCTGATGCGACGCTGCGCGTCTTATACTCCACACA TATGCCAGATTGAGCAACGGATACGGCTTCCCCAAGTTGCCCACTTCCATACGTGTCCTCCTTACCA GAAATTTATCCTTAAGATCGTTTAAACTCGACTCTGGCTCTATCGAATCTCCGTCGTTTCGAGCTTA CGCGAACAGCCGTGGCGCTCATTGCTCGTCGGGCATCGAATCTCGTCAGCTATCGTCAGCTTACCT TTTTGGCAGCGATCGCGGCTCCCGACATCTTGGACCATTAGCTCCACAGGTATCTTCTTCCCTCTAG TGGTCATAACAGCAGCTTCAGCTACCTCTCAATTCAAAAAACCCCTCAAGACCCGTTTAGAGGCCCC AAGGGGTTATGCTATCAATCGTTGCGTTACACACACAAAAACCAACACACATCCATCTTCGATGGA TAGCGATTTTATTATCTAACTGCTGATCGAGTGTAGCCAGATCTAGTAATCAATTACGGGGTCATTA GTTTCATAGCCCATATATGGAGTTCGCGGTTACATAACTTACGGTAAATGGCCCCGCTGGCTGACCGC CCAACGACCCCCGCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTT CCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCAT ATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCCGCTGGCATATGCCCAGTACA TGACCTTATGGGACTTTTCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGCTGAT GCGGTTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGAATCACGGGGATTTCGAAGTCTCCAC CCCATTGACGTCAATGGGAGTTTGTGTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTGCTAACA ACTCCGCCCCATTGACGCAAAATGGGCGGTACGCGGTACCGTGGGAGGTCTATATAAGCAGAGCTGG TTTAGTGAACCGTCAGATCAGATCTTTGTGTCAGTCTTACCATCCACTCGACACACCCGCGAGCGCGG CTAATACGACTCACTATAGGGAGAAGTACTGCCACCATGGGCAAGAAGCAAAATCGCAAGACGGGGGA ATTCCAAGACACAATCCGCTAGCCCACCACCTAAAGAGCGTTCTAGCTCCCCCTGCTACTGAGCAGTC CTGGATGGAACGACTTTCGATGAACTCCGGGAAGAGGGATTTAGGCGATCCAACCTATTCAGAACTC

	CTATCGCCTTCTTGACGAGTTCTTCTAGTATGTAAGCCCTGTGCCTTCTAGTTGCCAGCCATCTGTT GTTTGGCCCTCCCCCGTGCCTTTCCTTGACCCTGGAAGGTGCCACTCCCCTGTCCCTTTCCCTAATAAAA ATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCAATTCTATTCTGGGGGGTGGGGTGGGGCAGGA CAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGTTAAT TAACCAAGTCAAGTCAGCTACTTGGCGAGATCGACTTGTCTGGGTTTCGACTACGCTCAGAATTGCGT CAGTCAAGTTCGATCTGGTCCTTGCTATTGCACCCGTTCTCCGATTACGAGTTTCATTTAAATCATG TGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGC TCCGCCCCCTGACGAGCATCACAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACT ATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTT ACCGGATACCTGTCCGCTTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGT ATCTCAGTTCGGTGTAGGTGCTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCCGA CCGCTGCGCCTTATCCGGTAACCTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTG GCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAAGT GGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTAC CTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTT GTTTGCAAGCAGCAGATTACGCGCAGAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGG GGTCTGACGCTCAGTGAACGAAAACCTCACGTTAAGGGATTTTGGTCAAGAGATTATCAAAAAGGAT CTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTAAATCAATCTAAAGTATATATGAGTAACT TGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTTCGTTTCAT CCATAGTTGCATTTAAATTTCCGAACCTCTCCAAGGCCCTCGTCGGAAAAATCTTCAAACCTTTTCGTCC GATCCATCTTGCAGGCTACCTCTCGAACGAACCTATCGCAAGTCTCTTGGCCGGCCTTGCGCCTTGGC TATTGCTTGGCAGCGCCTATCGCCAGGTATTACTCCAATCCCGAATATCCGAGATCGGGATCACCCG AGAGAAGTTCAACCTACATCCTCAATCCCGATCTATCCGAGATCCGAGGAATATCGAAATCGGGGCG CGCCTGGTGTACCGAGAACGATCCTCTCAGTGCGAGTCTC (SEQ ID NO: 102)
RET-013 LINE1-GFP ORF2-N Nucleoplasmin NLS Linker	AGATCCGAGGAATATCGAAATCGGGGCGCGCCTGGTGTACCGAGAACGATCCTCTCAGTGCGAGTCT CGACGATCCATATCGTTGCTTGGCAGTCAGCCAGTCGGAATCCAGCTTGGGACCCAGGAAGTCCAAT GCTCAGATATTGTACTCAAGCCTGGTCAAGCGCAGCGCATCCGATCTGTTTAAACCTAGATATTGATAG TCTGATCGGTCAACGTATAATCGAGTCTTAGCTTTTGCAAACATCTATCAAGAGACAGGATCAGCAG GAGGCTTTTCGCATGAGTATTCAACATTTCCGTGTGCGCCTTATTCCCTTTTTTTGCGGCATTTTGCCT TCCTGTTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCGCGA GTGGGTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGCT TTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACGCCGGGCA AGAGCAACTCGGTGCGCGCATACTACTATTCTCAGAATGACTTGGTTGAGTATTCACCAGTCACAGAA AAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGAGTGATAACA CTGCGGCCAACTTACTTCTGACAACGATTGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACAT GGGGGATCATGTAACCTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAG CGTGACACCACGATGCCTGTAGCAATGGCAACAACCTTGGCTAAACTATTAAGTGGCGAACTACTTA CTCTAGCTTTCCCGCAACAGTTGATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCG CTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGT ATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTC AGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTA ACCGATTCTAGGTGCATTGGCGCAGAAAAAATGCCTGATGCGACGCTGCGCGTCTTATACTCCAC ATATGCCAGATTACAGAACGGATACGGCTTCCCCAAGTGGCCACTTCCATACGCTGTCTCTCTTACC AGAAATTTATCCTTAAGATCGTTTAAACTCGACTCTGGCTCTATCGAATCTCCGTCGTTTCGAGCTT ACGCGAACAGCCGTGGCGCTCATTTTGCTCGTGGGCATCGAATCTCGTCAGTCTCGTCAGCTTACC TTTTTGGCAGCGATCGCGGCTCCCGACATCTTGGACCATTAGCTCCACAGGTATCTTCTTCCCTCTA GTGGTCATAACAGCAGCTTCAGCTACCTCTCAATTCAAAAAACCCCTCAAGACCCGTTTAGAGGCCC CAAGGGGTTATGCTATCAATCGTTGCGTTACACACACAAAAACCAACACACATCCATCTTCGATGG ATAGCGATTTTATTATCTAACTGCTGATCGAGTGTAGCCAGATCTAGTAATCAATTACGGGGTCATT AGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCTGGCTGACCG CCCAACGACCCCCGCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTT TCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCA TATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCTGGCATTATGCCAGTAC ATGACCTTATGGGACTTTTCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGCTGA TGCGGTTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGAATCACGGGGGATTTCCAAGTCTCCA CCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAATAATGTCGTAAC AACTCCGCCCCATTGACGCAAAATGGGCGGTAGGCGGTGACGGTGGGAGGTCTATATAAGCAGAGCTG GTTTAGTGAACCGTCAGATCAGATCTTTGTGATCCTACCATCCACTCGACACACCCGCCAGCGGCC GCTAATACGACTCACTATAGGGAGAAGTACTGCCACCATGGGCAAGAAGCAAAATCGCAAGACGGGG AATTCGAAGACACAATCCGCTAGCCACCACTAAAGAGCGTTCTAGCTCCCCCTGCTACTGAGCAGT CCTGGATGGAAAAACGACTTCGATGAACCTCCGGGAAGGAGGATTTAGGCGATCCCAACTATTCAGA CCGCGAAGATATCCAGACAAAGGGGAAGGAAGTCGAGAATTTTCGAGAAGAACCTCAGAGAGTGATC ACCCGTATCACAAACACTGAGAAATGTCTCAAAGAAGTCAAGAACTTAAGACAAAAGCCAGGGAGC TTCGAGAGGAGTGTGCGAGTCTGAGATCCAGGTGTGACCAGCTCGAGGAGCGCGTGAGCGCGATGGA

	GGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGTTAATTAACCAAGTCAAGTCAGCTACTTTGGCGAGATCGACTTGTCTGGGTTTTCGACTACGCTCAGAATTGCGTCAGTCAAGTTCGATCTGGTCCTTGCTATTGCACCCGTTCTCCGATTACGAGTTTTCATTTAAATCATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCTGTTCGACCCTGCCGCTTACCGGATACCTGTCCGCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTGTTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCGTTTCAGACCCGACCGCTGCGCCTTATCCGGTAACATATCGTCTTGAGTCCAACCCGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTTCTACGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAATAATGAAGTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTTCGTTTCATCCATAGTTGCATTTAAATTTCCGAACCTCTCCAAGGCCCTCGTCGGAAAAATCTTCAAACCTTTTCGTCGGATCCATCTTGCAGGCTACCTCTCGAACGAACATATCGCAAGTCTCTTGGCCGGCCTTGCGCCTTGGCTATTGCTTGGCAGCGCCTATCGCCAGGTATTACTCCAATCCCGAATATCCGAGATCGGGATCACCCGAGAGAAGTTCAACCTACATCCTCAATCCCGATCTATCCG (SEQ ID NO:103)
RET-014 LINE1-GFP ORF2-C SV40-NLS	AGAACGATCCTCTCAGTGCGAGTCTCGACGATCCATATCGTTGCTTGGCAGTCAGCCAGTCGGAATCAGCTTGGGACCCAGGAAGTCCAATCGTCAGATATTGTAAGCTTGGTCACGGCAGCGTACCGATCTGTTTTAAACCTAGATATTGATAGTCTGATCGGTCAACGTATAATCGAGTCTTAGCTTTTGCAAAACATCTATCAAGAGACAGGATCAGCAGGAGGCTTTCGCATGAGTATTCAACATTTCCGTGTCGCCCTTATCCCTTTTTTGGCGCATTTTGCCTTCTGTTTTTGTCTACCCAGAAACGCTGGTGAAAGTAAAAAGATCTGAAGATCAGTTGGGTGCGCGAGTGGGTTTACATCGAAGTGGATCTCAACAGCGGTAAAGATCCTTGAGAGTTTTTCGCCCCGAAGAACGCTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGCGCGGTATTATCCCGTATTGACGCCGGGCAAGAGCAACTCGGTGCGCCGATACACTATTCTCAGAATGACTTGGTTGAGTATTACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGAGTGATAAACAAGTGGGCCAACTTACTTCTGACAACGATTGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAAGTTCGCTTGCATGCTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCAGATGCCTGTAGCAATGGCAACAACCTTGCCTAACTATTAAGTGGCGAAGTACTTACTCTAGCTTCCCGGCAACAGTTGATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCTCACTGATTAAGCATTGGTAACCGATTCTAGGTGCATTGGCGCAGAAAAAATGCCTGATGCGACGCTGCGCGTCTTATACTCCCACATATGCCAGATTTCAGCAACGGATACGGCTTCCCCAACTTGCCCACTTCCATACGTGTCTCCTTACCAGAAATTTATCCTTAAGATCGTTTTAAACTCGACTCTGGCTCTATCGAATCTCCGTGTTTTCGAGCTTACGCGAACAGCCGTGGCGCTCATTTGCTCGTCGGGCATCGAATCTCGTCAGCTATCGTCAGCTTACCTTTTTTGGCAGCGATCGCGCTCCCGACATCTTGGACCATTAGCTCCACAGGTATCTTCTTCCCTCTAGTGGTCATAACAGCAGCTTCAGCTACCTCTCAATTCAAAAAACCCCTCAAGACCCGTTTAGAGGCCCAAGGGGTTATGCTATCAATCGTTGCGTTACACACACAAAAACCAACACACATCCATCTTTCGATGGATAGCGATTTTATTATCTAAGTCTGCTGAGTGTAGCCAGATCTAGTAATCAATTACGGGGTCATTAGTTTCATAGCCCATATATGGAGTTCCGCGTTACATAAATTACGGTAATGGCCCGCTGGCTGACCGCCCAACGACCCCCGCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCCTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCGCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGCTGATGCGTTTTTGGCAGTACATCAATGGGCGTGATAGCGGTTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTTGACGTCAATGGGAGTTTGTGTTTGGCACCAAAATCAACGGGACTTTCCAATGTTCGTAACAACCTCCGCCCCATTGACGCAAAATGGGCGGTAGGCGGTGACGGTGGGAGGTCTATATAAGCAGAGCTGGTTTTAGTGAACCGTCAGATCAGATCTTTGTTCGATCTTACCATCCACTCGACACACCCGCCAGCGGCCGCTAATACGACTCACTATAGGGAGAAAGTACTGCCACCATGGGCAAGAAGCAAAATCGCAAGACGGGAATTTCCAAGACACAATCCGCTAGCCCCACCACCTAAAGAGCGTTCTAGCTCCCCTGCTACTGAGCAGTCTGGATGGAAAACGACTTCGATGAACTCCGGGAAGAGGGATTTAGGCGATCCAATATTGAGAACTCCGCGAAGATATCCAGACAAAGGGGAAGGAAGTCGAGAATTTGAGAAAGAACTCGAGGAGTGCATACCCGTATCACAAACACTGAGAAATGTCTCAAAGAACTCATGGAACCTTAAGACAAAAGCCAGGGAGCTTCGAGAGGAGTGTGCGAGTCTGAGATCCAGGTGTGACCAGCTCGAGGACCTGAGCGCGATGGAAGACGAGATGAACGAGATGAAAAGAGAGGGCAAAATTCAGGGAAGAGCGCATTAAGAGGAACGAACAGAGTCTGAGGAGATTTGGGATTACGTCAAGAGGCCTAACCTGCGTTGATCGGCGTCCCCGAGAGCGACGTAGAAAACGGGACTAAACTGGAGAATACACTTCAAGACATCATTTCAAGAAAATTTTCCAAACCTGGCTCGGCAAGCTAATGTGCAAAATCCAAGAGATCCAACGCACACCCAGCGGTATAGCTCTCGGCGTGCCACCCCTAGGCATATTATCGTGCGCTTTACTAAGGTGGAG

	TAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCCGTGTAGGTGCTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTACAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTTCGTTTCATCATAGTTGCATTTTAAATTTCCGAACCTCTCCAAGGCCCTCGTCGGAAAAATCTTCAAACCTTTTCGTCCGATCCATCTTGCAGGCTACCTCTCGAACGAACCTATCGCAAGTCTCTTGGCCGGCCTTGCGCCTTGCGCTATTGCTTGGCAGCGCCTATCGCCAGGTATTACTCCAATCCCGAATATCCGAGATCGGGATCACCCGAGAGAAGTTCAACCTACATCCTCAATCCCGATCTATCCGAGATCCGAGGAATATCGAAATCGGGGCGCGCCTGGTGTACCG (SEQ ID NO: 104)
RET-015 LINE1-GFP ORF2-C Linke r_SV40-NLS	TCGACGATCCATATCGTTGCTTGGCAGTCAGCCAGTCGGAATCCAGCTTGGGACCCAGGAAGTCCAA TCGTGAGATATTGTACTCAAGCCTGGTCACGGCAGCGTACCGATCTGTTTAAACCTAGATATTGATA GTCTGATCGGTCAACGTATAATCGAGTCCTAGCTTTTGCAAACATCTATCAAGAGACAGGATCAGCA GGAGGCTTTTCGCATGAGTATTCAACATTTCCGTGTGCGCCCTTATTCCCTTTTTTTCGGCATTTTGCC TTCTGTTTTTGTCTACCCAGAAACGCTGGTGAAAGTAAAGATGCTGAAGATCAGTTGGGTGCGCG AGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTTCGCCCCGAAGAACGC TTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACGCCGGGC AAGAGCAACTCGGTCGCCGCATACACTATTCTCAGAATGACTTGGTTGAGTATTCACCAGTCACAGA AAAGCATCTTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGAGTGATAAC ACTGCGGCCAACTTACTTCTGACAACGATTGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACA TGGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATCAACCAACGACA GCGTGACACCACGATGCCTGTAGCAATGGCAACAACCTTGCCTAAACTATTAAGTGGCGAACTACTT ACTCTAGCTTCCCGGCAACAGTTGATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGC GCTCGGCCCTTCCGGCTGGCTGGTTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGG TATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGT CAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGT AACCGATTCTAGGTGCATTGGCGCAGAAAAAATGCCTGATGCGACGCTGCGCGTCTTATACTCCCA CATATGCCAGATTTCAGCAACGGATACGGCTTCCCCAACTTGCCCACTTCCATACGTGTCTCTCTTAC CAGAAATTTATCCTTAAGATCGTTTTAACTCGACTCTGGCTCTATCGAATCTCCGTGCTTTCGAGCT TACGCGAACAGCCGTGGCGCTCATTTGCTCGTGGGCATCGAATCTCGTCAGCTATCGTCAGCTTAC CTTTTTGGCAGCGATCGCGGCTCCCGACATCTTGGACCATTAGCTCCACAGGTATCTTCTTCCCTCT AGTGGTCATAACAGCAGCTTCAGCTACCTCTCAATTCAAAAAACCCCTCAAGACCCGTTTAGAGGCC CCAAGGGGTTATGCTATCAATCGTTGCGTTACACACACAAAAAACCAACACACATCCATCTTCGATG GATAGCGATTTTATTATCTAACTGCTGATCGAGTGTAGCCAGATCTAGTAATCAATTACGGGGTCAT TAGTTTCATAGCCCATATATGGAGTTCGCGGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACC GCCCCAACGACCCCCGCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACT TTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGATC ATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCAGTA CATGACCTTATGGGACTTTCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGCTG ATGCGGTTTTTGGCAGTACATCAATGGGCGTGATAGCGGTTTGAATCACGGGGATTTTCAAAGTCTCC ACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTTCAAATGTGCTAA CAACTCCGCCCATTTGACGCAATGGGCGGTAGGCGGTGACGGTGGGAGGTCTATATAAGCAGAGCT GGTTTAGTGAACCGTCAGATCAGATCTTTGTGATCCTACCATCCACTCGACACACCCGCCAGCGGC CGCTAATACGACTCACTATAGGGAGAAGTACTGCCACCATGGGCAAGAAGCAAAATCGCAAGACGGG GAATTCGAAGACACAATCCGCTAGCCACACCTAAAGAGCGTTCTAGCTCCCTGCTACTGAGCAG TCCTGGATGGAAAACGACTTCGATGAACTCCGGGAAGAGGGATTTAGGCGATCCAACCTATTCAGAAC TCCGCGAAGATATCCAGACAAAGGGGAAGGAAGTCGAGAATTTGAGAAGAACCTCGAGGAGTGCA CACCCGTATCACAACACTGAGAAATGTCTCAAAGAACTCATGGAACCTTAAGACAAAAGCCAGGGAG CTTCGAGAGGAGTGTGCGAGTCTGAGATCCAGGTGTGACCAGCTCGAGGAGCGCGTGAGCGCGATGG AAGACGAGATGAACGAGATGAAAAGAGAGGGCAAATTCAGGGAGAAGCGCATTAAGAGGAACGAACA GAGTCTGCAGGAGATTTGGGATTACGTCAAGAGGCCTAACCTGCGGTTGATCGGCGTCCCCGAGAGC GACGTAGAAAACGGGACTAACTGGAGAATACACTTCAAGACATCATTCAAGAAAATTTTCCAAACC TGGCTCGGCAAGCTAATGTGCAAATCCAAGAGATCCAACGCACACCCAGCGGTATAGCTCTCGGCG TGCCACCCCTTAGGCATATTATCGTGCCTTTTACTAAGGTGGAGATGAAAGAGAAGATGCTGCGAGCC GCTCGGGAAGGAAGGGTGACTTTGAAGGGCAAACCTATTTCGGCTGACGGTTGACCTTAGCGCCG AGACACTCCAGGCACGCCGGAATGGGGCCCCATCTTAAATATCCTGAAGGAGAAGAACTTCCAGCC ACGAATCTCTTACCCTGCAAAGTTGAGTTTTATCTCCGAGGGTGAGATTAAGTATTTTCATCGATAAA CAGATGCTGCGAGACTTCGTGACAACCTCGCCAGCTCTCAAGGAACTGCTCAAAGAGGCTCTTAATA

	GTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTA CGGCTACACTAGAGAAGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGA GTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTTGTTTGCAAGCAGC AGATTACGCGCAGAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCA GTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATC CTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAACTTGGTCTGACAGTT ACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTTCGTTTCATCCATAGTTGCATT TAAATTTCCGAACCTCTCCAAGGCCCTCGTCGGAAAAATCTTCAAACCTTTTCGTCCGATCCATCTTGCA GGCTACCTCTCGAACGAACATATCGCAAGTCTCTTGGCCGGCCTTGCGCCTTGGCTATTGCTTGGCAG CGCCTATCGCCAGGTATTACTCCAATCCCGAATATCCGAGATCGGGATCACCCGAGAGAAGTTCAAC CTACATCCTCAATCCCGATCTATCCGAGATCCGAGGAATATCGAAATCGGGGCGCGCCTGGTGTACC GAGAACGATCCTCTCAGTGCGAGTC (SEQ ID NO: 105)
ret-016- line1- gfp- orf2- c_nucleo plasmin- nls	ACCGAGAACGATCCTCTCAGTGCGAGTCTCGACGATCCATATCGTTGCTTGGCAGTCAGCCAGTCGG AATCCAGCTTGGGACCCAGGAAGTCCAATCGTCAGATATTGTACTCAAGCCTGGTCACGGCAGCGTA CCGATCTGTTTTAAACCTAGATATTGATAGTCTGATCGGTCAACGTATAATCGAGTCTTAGCTTTTGC AAACATCTATCAAGAGACAGGATCAGCAGGAGGCTTTTCGCATGAGTATTCAACATTTCCGTGTCGCC CTTATTCCCTTTTTTGGCGCATTTTGCCTTCCTGTTTTTGTCTACCCAGAAACGCTGGTGAAGTAA AAGATGCTGAAGATCAGTTGGGTGCGCGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGAT CCTTGAGAGTTTTTCGCCCCGAAGAACGCTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGC GCGGTATTATCCCGTATTGACGCCGGGCAAGAGCAACTCGGTGCGCCGATACACTATTCTCAGAATG ACTTGGTTGAGTATTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATG CAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATTGGAGGACCG AAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACCTCGCCTTGATCGTTGGGAACCGG AGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTGTAGCAATGGCAACAACCTT GCGTAAACTATTAAGTGGCGAACTACTTACTCTAGCTTCCCGGCAACAGTTGATAGACTGGATGGAG GCGGATAAAGTTGCAGGACCATTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTATTGCTGATAAAT CTGGAGCCGGTGAGCGTGGGTCTCGCGTATCATTCGACGACTGGGGCCAGTTAGTACGCTCCCG TATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAG ATAGGTGCCTCACTGATTAAGCATTGGTAACCGATTCTAGGTGCATTGGCGCAGAAAAAATGCCTG ATGCGACGCTGCGCGTCTTATACTCCACATATGCCAGATTGAGCAACGGATACGGCTTCCCCAACT TGCCCACTTCCATACGTGTCCTCCTTACCAGAAATTTATCCTTAAGATCGTTTAAACTCGACTCTGG CTCTATCGAATCTCCGTCGTTTCGAGCTTACGCGAACAGCCGTGGCGCTCATTGCTCGTCGGGCAT CGAATCTCGTCAGCTATCGTCAGCTTACCTTTTTTGGCAGCGATCGCGCTCCCGACATCTTGGACCA TTAGCTCCACAGGTATCTTCTTCCCTCTAGTGGTCATAACAGCAGCTTCAGCTACCTCTCAATTCAA AAAACCCCTCAAGACCCGTTTAGAGGCCCAAGGGGTTATGCTATCAATCGTTGCGTTACACACACA AAAAACCAACACACATCCATCTTCGATGGATAGCGATTTTATTATCTAACTGCTGATCGAGTGATAGC CAGATCTAGTAATCAATTACGGGGTCATTAGTTTCATAGCCCATATATGGAGTTCCGCGTTACATAAC TTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCATTGACGTCAATAATGACGTA TGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACT GCCCCTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTA AATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTA CGTATTAGTCATCGCTATTACCATGCTGATCGGTTTTTGGCAGTACATCAATGGGCGTGGATAGCGG TTTGACTACGGGGATTTCCAAGTCTCCACCCCATTTGACGTCATGGGAGTTTGTTTTGGCACCAAA ATCAACGGGACTTTCCAAGTGTGTAACAACCTCGCCCCATTGACGCAAAATGGGCGGTAGGCGTGT ACGGTGGGAGGTCTATATAAGCAGAGCTGGTTTAGTGAACCGTCAGATCAGATCTTTGTCGATCCTA CCATCCACTCGACACACCCGCCAGCGGCCGCTAATACGACTCACTATAGGGAGAAGTACTGCCACCA TGGGCAAGAAGCAAAATCGCAAGACGGGGAATTCCAAGACACAATCCGCTAGCCCACCACCTAAAGA GCGTTCTAGCTCCCCTGCTACTGAGCAGTCTGGATGGAAAACGACTTCGATGAACTCCGGGAAGAG GGATTTAGGCGATCCAATATTGAGAACTCCGCGAAGATATCCAGACAAAGGGGAAGGAAGTCGAGA ATTTGAGAAGAACCTCGAGGAGTGCATCACCCGTATCACAAACACTGAGAAATGTCTCAAAGAAGT CATGGAACCTTAAGACAAAAGCCAGGGAGCTTCGAGAGGAGTGTGCGAGTCTGAGATCCAGGTGTGAC CAGCTCGAGGAGCGCGTGAGCGCGATGGAAGACGAGATGAACGAGATGAAAAGAGAGGGCAAATTC GGGAGAAGCGCATTAAGAGGAACGAACAGAGTCTGCAGGAGATTTGGGATTACGTCAAGAGGCCTAA CCTGCGGTTGATCGGCGTCCCCGAGAGCGACGTAGAAAACGGGACTAACTGGAGAATACACTTCAA GACATCATTCAAGAAAATTTTCCAACCTGGCTCGGCAAGCTAATGTGCAAAATCCAAGAGATCCAAC GCACACCCAGCGGTATAGCTCTCGGCGTGCCACCCCTAGGCATATTATCGTGCGCTTTACTAAGGT GGAGATGAAAGAGAAGATGCTGCGAGCCGCTCGGGAAAAGGGAAGGGTGACTTTGAAGGGCAAACCT ATTGCGCTGACGGTTGACCTTAGCGCCGAGACACTCCAGGCACGCCGGGAATGGGGCCCCATCTTTA ATATCCTGAAGGAGAAGAACTTCCAGCCACGAATCTCTTACCCTGCAAAGTTGAGTTTATCTCCGA GGGTGAGATTAGTATTTTCATCGATAAACAGATCTCGCAGACTTCGTGACAACCTCGCCAGCTCTC AAGGAAGTCTCAAAGAGGCTCTTAATATGAGAGCGCAATAATAGATATCAACCTTGCAGAACCACG CAAAGATGTGAGACAGCCGTACAGCATCAAGACTAGGAAGAACTGCATCAACTAATGAGCAAAAT CACCAGCTAACATCATAGTATACATGACCGGCTCTAACTCACATATCACCATCCTTACACTTAACAT TAACGGCCTCAACTCAGCTATCAAGCGCCATCGGCTGGCCAGCTGGATCAAATCACAGGATCCAAGC

	CTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAAACGAAAACCTCACGTTAAGG GATTTTGGTCATGAGATTATCAAAAAGGATCTTTCACCTAGATCCTTTTAAATTAATAAATGAAGTTTT AAATCAATCTAAAGTATATATGAGTAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCAC CTATCTCAGCGATCTGTCTATTTTCGTTTCATCCATAGTTGCATTTAAATTTCCGAACTCTCCAAGGCC CTCGTCGGAAAACTTTCAAACCTTTTCGTCGATCCATCTTGACAGGCTACCTCTCGAACGAACTATCG CAAGTCTCTTGGCCGGCCTTGCCTTGGCTATTGCTTGGCAGCGCCTATCGCCAGGTATTACTCCA ATCCCGAATATCCGAGATCGGGATCACCCGAGAGAAGTTCAACCTACATCTCAATCCCGATCTATC CGAGATCCGAGGAATATCGAAATCGGGGCGCGCTGGTGT (SEQ ID NO: 106)
RET- 017 LINE1- GFP ORF2- C_Linke r_Nucle oplasm n-NLS	CCGAGAACGATCCTCTCAGTGCGAGTCTCGACGATCCATATCGTTGCTTGGCAGTCAGCCAGTCGGA ATCCAGCTTGGGACCCAGGAAGTCCAATCGTCAGATATTGTACTCAAGCCTGGTCACGGCAGCGTAC CGATCTGTTTAAACCTAGATATTGATAGTCTGATCGGTCAACGTATAATCGAGTCCTAGCTTTTGCA AACATCTATCAAGAGACAGGATCAGCAGGAGGCTTTTCGATGAGTATTCAACATTTCCGTGTCGCCC TTATTCCCTTTTTTTCGCGCATTTTTCCTTCTGTTTTTTCGCTCACCCAGAAACGCTGGTGAAAGTAAA AGATGCTGAAGATCAGTTGGGTGCGCGAGTGGGTACATCGAACTGGATCTCAACAGCGGTAAAGATC CTTGAGAGTTTTTCGCCCCGAAGAACGCTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCG CGGTATTATCCCGTATTGACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCTCAGAATGA CTTGGTTGAGTATTACACAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGC AGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATTGGAGGACCGA AGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACCTCGCCTTGATCGTTGGGAACCGGA GCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTGTAGCAATGGCAACAACCTTG CGTAAACTATTAACCTGGCGAACTACTTACTCTAGCTTCCCGGCAACAGTTGATAGACTGGATGGAGG CGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATC TGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGT ATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGA TAGGTGCCTCACTGATTAAGCATTGGTAACCGATTCTAGGTGCATTGGCGCAGAAAAAATGCCTGA TGCGACGCTGCGCGTCTTATACTCCCACATATGCCAGATTTCAGCAACGGAATACGGCTTCCCCA GCCCCACTTCCATACGTGTCCTCCTTACCAGAAATTTATCCTTAAGATCGTTTAAATCTGCTTGGC TCTATCGAATCTCCGTCTGTTTCGAGCTTACGCGAACAGCCGTGGCGCTCATTGCTCGTCGGGCATC GAATCTCGTCAGCTATCGTCAGCTTACCTTTTTTGGCAGCGATCGCGGCTCCCGACATCTTGGACCAT TAGCTCCACAGGTATCTTCTTCCCTCTAGTGGTCATAACAGCAGCTTCAGCTACCTCTCAATTCAAA AAACCCCTCAAGACCCGTTTAGAGGCCCAAGGGGTATGCTATCAATCGTTGCGTTACACACACAA AAAACCAACACACATCCATCTTCGATGGATAGCGATTTTATTATCTAACTGCTGATCGAGTGATGCC AGATCTAGTAATCAATTACGGGGTCATTAGTTTCATAGCCCATATATGGAGTTCGCGTTACATAACT TACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTAT GTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTG CCCACTTGGCAGTACATCAAGTGATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAA ATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTTCTACTTGGCAGTACATCTAC GTATTAGTCATCGCTATTACCATGCTGATGCGGTTTTTGGCAGTACATCAATGGGCGTGGATAGCGGT TTGACTCACGGGGATTTCCAAGTCTCCACCCATTGACGTCAATGGGAGTTTGTGTTTGGCACCAAAA TCAACGGGACTTTCCAAAATGTCGTAACAACCTCCGCCCATTTGACGCAAATGGGCGGTAGGCGTGTA CGGTGGGAGGTCTATATAAGCAGAGCTGGTTTAGTGAAACCGTCAGATCAGATCTTTGTCGATCCTAC CATCCACTCGACACACCCGCCAGCGGCCGCTAATACGACTCACTATAGGGAGAAGTACTGCCACCAT GGGCAAGAAGCAAAATCGCAAGACGGGAATTCGAAGACACAATCCGCTAGCCCCACCACTAAAGAG CGTTCTAGTCTCCCTGCTACTGAGCAGTCTGGATGGAAAACGACTTCGATGAATCCGGGAAGAGG GATTTAGGCGATCCAATATTGAGAACTCCGCGAAGATATCCAGACAAAGGGGAAGGAAGTCGAGAA TTTTGAGAAGAACCTCGAGGAGTGCATCACCCGTATCACAACACTGAGAAATGTCTCAAAGAACTC ATGGAACCTTAAGACAAAAGCCAGGGAGCTTCGAGAGGAGTGTGCGAGTCTGAGATCCAGGTGTGACC AGCTCGAGGAGCGCGTGAGCGCGATGGAAGACGAGATGAACGAGATGAAAAGAGAGGGCAAAATTCAG GGAGAAGCGCATTAAAGAGGAACGAACAGAGTCTGCAGGAGATTTGGGATTACGTCAAGAGGCCCTAAC CTGCGGTTGATCGGCGTCCCCGAGAGCGACGTAGAAAACGGGACTAACTGGAGAATACACTTCAAG ACATCATTCAAGAAAATTTTCCAAACCTGGCTCGGCAAGCTAATGTGCAAAATCCAAGAGATCCAACG CACACCCAGCGGTATAGCTCTCGGCGTGCCACCCCTAGGCATATTATCGTGCGCTTTACTAAGGTG GAGATGAAAGAGAAGATGCTGCGAGCCGCTCGGGAAAAGGGAAGGGTGACTTTGAAGGGCAAACCTA TTCGGCTGACGGTTGACCTTAGCGCCGAGACACTCCAGGCACGCCGGGAATGGGGCCCCATCTTTAA TATCCTGAAGGAGAAGAATTCAGCCACGAATCTCTTACCCTGCAAAGTTGAGTTTTATCTCCGAG GGTGAGATTAAGTATTTTCATCGATAAACAGATGCTGCGAGACTTCGTGACAACTCGCCCAGCTCTCA AGGAACCTGCTCAAAGAGGCTCTTAATATGGAGCGCAATAATAGATATCAACCTTGCAGAACACGC AAAGATGTGAGACAGCCGTGAGACCATCAAGACTAGGAAGAACTGCATCAACTAATGAGCAAAATC ACCAGCTAACATCATAGTATACATGACCGGCTCTAACTCACATATACCATCTTACTTAACTTAACTT AACGGCCTCAACTCAGCTATCAAGCGCCATCTGGCTGGCCAGCTGGATCAAATCACAGGATCCAAGCG TTTGTGTCATCCAAGAGACCCACCTGACCTGTAGAGATACTACCCGCTCAAGATCAAGGGATGGCG AAAGATTTATCAGGCGAACGGTAAGCAGAAGAAAGCCGGAGTCGAATCTGGTCTCAGACAAGACG GATTTCAAGCCCAAAAATTAAGCGTGATAAGGAAGGTCACTATATTATGGTGAAAGGCAGCATAAC AGCAGGAAGAATTAACATATTGAACATCTACGCGCCAAACACCGGCGCACCTCGCTTTATCAAAACA

	TCTCCAAGGCCCTCGTCGGAAAAATCTTCAAACCTTTTCGTCCGATCCATCTTGCAGGCTACCTCTCGA ACGAACTATCGCAAGTCTCTTGGCCGGCCTTGGCCCTTGGCTATTGCTTGGCAGCGCCTATCGCCAG GTATTACTCCAATCCCGAATATCCGAGATCGGGATCACCCGAGAGAAGTTTCAAACCTACATCCTCAAT CCCGATCTATCCGAGATCCGAGGAATATCGAAATCGGGGCGCGCCTGGTGTA (SEQ ID NO:107)
RET- 018 LINE1- GFP ORF2- N_SV40 _NLS_L inker ORF2- C_Nucle oplasm n_NLS	TGGCCGGCCTTGGCCCTTGGCTATTGCTTGGCAGCGCCTATCGCCAGGTATTACTCCAATCCCGAAT ATCCGAGATCGGGATCACCCGAGAGAAGTTTCAAACCTACATCCTCAATCCCGATCTATCCGAGATCCG AGGAATATCGAAATCGGGGCGCGCCTGGTGTACCGAGAACGATCCTCTCAGTGCGAGTCTCGACGAT CCATATCGTTGCTTGGCAGTCAGCCAGTCGGAATCCAGCTTGGGACCCAGGAAGTCCAATCGTCAGA TATTGTACTCAAGCCTGGTCACGGCAGCGTACCGATCTGTTTAAACCTAGATATTGATAGTCTGATC GGTCAACGTATAATCGAGTCCTAGCTTTTGCAAACATCTATCAAGAGACAGGATCAGCAGGAGGCTT TCGCATGAGTATTCAACATTTCCGTGTGCGCCCTTATTCCCTTTTTTGGCGGCATTTTGCCCTTCTGT TTTGCTCACCCAGAAACGCTGGTGAAAGTAAAGATGCTGAAGATCAGTTGGGTGCGCGAGTGGGT ACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGCTTTCCAAT GATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACGCCGGGCAAGAGCAA CTCGGTGCGCGCATACACTATTCTCAGAATGACTTGGTTGAGTATTACCAGTCACAGAAAAGCATC TTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGC CAACTTACTTCTGACAACGATTGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGAT CATGTAACCTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACA CCACGATGCCTGTAGCAATGGCAACAACCTTGCCTAAACTATTAAGTGGCGAACTACTTACTCTAGC TTCCCGGCAACAGTTGATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCC CTTCCGGCTGGCTGGTTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTG CAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAAC TATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTTGGTAACCGATT CTAGGTGCATTGGCGCAGAAAAAATGCCTGATGCGACGCTGCGCGTCTTATACTCCACATATGCC AGATTACGAAACGGATACGGCTTCCCCAAGTTGCCCACTTCCATACGTGTCTCTTACCAGAAAT TATCCTTAAGATCGTTTAAACTCGACTCTGGCTATCTGAATCTCCGTCGCTTTCGAGCTTACGCGAA CAGCCGTGGCGCTCATTTGCTCGTCGGGCATCGAATCTCGTCAGCTATCGTCAGCTTACCTTTTGG CAGCGATCGCGCTCCCGACATCTTGGACCATTAGCTCCACAGGTATCTTCTTCCCTCTAGTGGTCA TAACAGCAGCTTCAGCTACCTCTCAATTCAAAAAACCCCTCAAGACCCGTTTAGAGGCCCCAAGGGG TTATGCTATCAATCGTTGCGTTACACACACAAAAACCAACACACATCCATCTTCGATGGATAGCGA TTTTATTATCTAACTGCTGATCGAGTGTAGCCAGATCTAGTAATCAATTACGGGGTCATTAGTTTCA AGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACG ACCCCGCCCATTTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTG ACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGATCATATGCCA AGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCAGTACATGACCT TATGGGACTTTCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGCTGATGCGGTT TTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCAT GACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCG CCCCATTGACGCAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTGGTTTAGT GAACCGTCAGATCAGATCTTTGTGATCCTACCATCCACTCGACACACCCGCCAGCGCCGCTAATA CGACTCACTATAGGGAGAAGTACTGCCACCATGGGCAAGAAGCAAAATCGCAAGACGGGGAAATCCA AGACACAATCCGCTAGCCCACCACCTAAAGAGCGTTCTAGCTCCCCTGCTACTGAGCAGTCTTGAT GGAAAACGACTTCGATGAACTCCGGGAAGAGGATTTAGGCGATCCAATATTGAGAACTCCGCGAA GATATCCAGACAAAGGGGAAGGAAGTCGAGAATTTGAGAAGAACCTCGAGGAGTGATCACCCTGTA TCACAAACACTGAGAAATGTCTCAAAGAAGTCAATGGAACCTAAGACAAAAGCCAGGGAGCTTCGAGA GGAGTGTGCGAGTCTGAGATCCAGGTGTGACCAGCTCGAGGAGCGCGTGAGCGCGATGGAAGACGAG ATGAACGAGATGAAAAGAGAGGGGCAAAATTCAGGGAGAAGCGCATTAAGAGGAACGAACAGAGTCTGC AGGAGATTTGGGATTACGTCAAGAGGCCTAACCTGCGGTTGATCGGCGTCCCCGAGAGCGACGTAGA AAACGGGACTAACTGGAGAATACACTTCAAGACATCATTCAAGAAAATTTTCAAACCTGGCTCGG CAAGCTAATGTGCAAATCCAAGAGATCCAACGCACACCCAGCGGTATAGCTCTCGGCGTGCCACCC CTAGGCATATTATCGTGCCTTTACTAAGGTGGAGATGAAAGAGAAGATGCTGCGAGCCGCTCGGGA AAAGGGAAGGGTGACTTTGAAGGGCAAACCTATTGCGCTGACGGTTGACCTTAGCGCCGAGACACTC CAGGCACGCCGGGAATGGGGCCCCATCTTTAATATCCTGAAGGAGAAGAACTTCCAGCCACGAATCT CTTACCCTGCAAAGTTGAGTTTTATCTCCGAGGGTGAGATTAAGTATTTTCATCGATAAACAGATGCT GCGAGACTTCGTGACAACCTCGCCAGCTCTCAAGGAACTGCTCAAAGAGGCTCTTAATATGGAGCGC AATAATAGATATCAACCCTTGAGAAACACGCAAGATGTGAGACAGCCGTGAGACCATCAAGACTA GGAAGAACTGCATCAACTAATGAGCAAAATCACAGCTAACATCATAGTATACATGccaaagaaga agcggaaggtcggcgggcggcagcACCGGCTCTAACTCACATATCACCATCCTTACACTTAACATTAA CGGCCTCAACTCAGCTATCAAGCGCCATCGGCTGGCCAGCTGGATCAAATCACAGGATCCAAGCGTT TGTTGCATCCAAGAGACCCACCTGACCTGTAGAGTACTCACCAGCTCAAGGATCAAGGGATGGCGAA AGATTTATCAGGCGAACGGTAAGCAGAAGAAAGCCGAGTCGCAATTTCTGGTCTCAGACAAGACGGA TTTCAAGCCACCAAAATTAAGCGTGATAAGGAAGGTCACTATATTATGGTGAAAGGCAGCATACAG CAGGAAGAACTTACCATATTGAACATCTACGCGCCAAACACCGGCGCACCTCGCTTTATCAAACAGG TCCTGTCCGATCTGCAGCGAGATCTGGATTCTCATACGTTGATTATGGGTGATTTCAATACACCATT

	GTCGGAAAATCTTCAAACCTTTTCGTCCGATCCATCTTGCAGGCTACCTCTCGAACGAACTATCGCAA GTCTCT (SEQ ID NO: 108)
ret-028-line1-orf1-t2a-orf2-n_sv40_nls_linker-gfp	TGGCCGGCCTTGGCGCTTGGCTATTGCTTGGCAGCGCCTATCGCCAGGTATTACTCCAATCCCCGAAT ATCCGAGATCGGGATCACCCGAGAGAAGTTCAACCTACATCCTCAATCCCGATCTATCCGAGATCCG AGGAATATCGAAATCGGGGCGCGCCTGGTGTACCGAGAACGATCCTCTCAGTGCGAGTCTCGACGAT CCATATCGTTTGCTTGGCAGTCAGCCAGTCGGAATCCAGCTTGGGACCCAGGAAGTCCAATCGTCAGA TATTGTACTCAAGCCTGGTCACGGCAGCGTACCGATCTGTTTAAACCTAGATATTGATAGTCTGATC GGTCAACGTATAATCGAGTCTTAGCTTTTGCAAACATCTATCAAGAGACAGGATCAGCAGGAGGCTT TCGCATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCTTTTTTGCGGCATTTTGCCTTCCTGTT TTTGCTCACCCAGAAACGCTGGTGAAAGTAAAGATGCTGAAGATCAGTTGGGTGCGCGAGTGGGTT ACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTTCGCCCCGAAGAACGCTTTCCAAT GATGAGCACTTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACGCCGGGCAAGAGCAA CTCGGTGCCTGCATACACTATTCTCAGAATGACTTGGTTGAGTATTCACCAGTCACAGAAAAGCATC TTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGC CAACTTACTTCTGACAACGATTGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGAT CATGTAACCTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATAACAAACGACGAGCGTGACA CCACGATGCCTGTAGCAATGGCAACAACCTTGCCTAACTATTAAGTGGCGAACTACTTACTCTAGC TTCCCGGCAACAGTTGATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCC CTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTG CAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAAC TATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAACCGATT CTAGGTGCATTGGCGCAGAAAAAATGCCTGATGCGACGCTGCGCGTCTTATACTCCACATATGCC AGATTGAGCAACGGATACGGCTTCCCCAAGTTCCTGACCTTCCATACGTGTCTCTTACCAGAAAT TATCCTTAAGATCGTTTAACTCGACTCTGGCTCTATCGAATCTCCGTGCTTTCGAGCTTACGCGAA CAGCGTGGCGCTCATTTGCTCGTGGGCATCGAATCTCGTCAGCTATCGTCAGCTTACCTTTTTTG CAGCGATCGCGCTCCCGACATCTTGGACATTAGCTCCACAGGTATCTTCTCCCTCTAGTGGTCA TAACAGCAGCTTCAGCTACCTCTCAATCAAAAAACCCCTCAAGACCCGTTTAGAGCCGCAAGGGG TTATGCTATCAATCGTTGCGTTACACACAAAAAACCAACACACATCCATTCTGATGGATAGCGA TTTTATTATCTAAGTCTGATCGAGTGTAGCCAGATCTAGTAATCAATTACGGGGTCATTAGTTTCAT AGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCCGCTGGCTGACCGCCCAACG ACCCCGCCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTG ACGTCAATGGGTGGAGTATTTACGGTAACTGCCACTTGGCAGTACATCAAGTGTATCATATGCCA AGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCCGCTGGCATTATGCCAGTACATGACCT TATGGGACTTTCTTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGCTGATGCGGTT TTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGAAGTACGCGGGGATTTCGAAGTCTCCACCCATT GACGTCAATGGGAGTTTGTGTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACCTCCG CCCCATTGACGCAAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTGGTTTAGT GAACCGTCAGATCAGATCTTTGTGATCCTACCATCCACTCGACACACCCGCCAGCGGCCGCTAATA CGACTCACTATAGGGAGAAGTACTGCCACCATGGGCAAGAAGCAAAATCGCAAGACGGGGGAATTCCA AGACACAATCCGCTAGCCCACCACCTAAAGAGCGTTCTAGCTCCCCTGCTACTGAGCAGTCTTGAT GGAAAACGACTTCGATGAACTCCGGGAAGAGGGATTTAGGCGATCCAATATTCAGAACTCCGCGAA GATATCCAGACAAAGGGGAAGGAAGTCGAGAATTTGAGAAGAACCTCGAGGAGTGCATCACCCGTA TCACAACTCAGTGAAGATGTCTCAAAGAATCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA AGAGTTCGAGTCTGAGATCCAGGTGTGACCAGTCTGAGGAGCGCGTGAAGCAGATGGAAGACGAG ATGAACGAGATGAAAAGAGAGGGCAAAATTCAGGGAGAAGCGCATTAAGAGGAACGAAACAGAGTCTGC AGGAGATTTGGGATTACGTCAAGAGGCCTAACCTGCGGTTGATCGGCGTCCCCGAGAGCGACGTAGA AAACGGGACTAACTGGAGAATACACTTCAAGACATCATTCAAGAAAAATTTTCCAAACCTGGCTCGG CAAGCTAATGTGCAAAATCCAAGAGATCCAACGCACACCCAGCGGTATAGCTCTCGGCGTGCCACCC CTAGGCATATTATCGTGCGCTTTACTAAGGTGGAGATGAAAGAGAAGATGCTGCGAGCCGCTCGGGA AAAGGGAAGGGTGACTTTGAAGGGCAAACCTATTTCGGCTGACGGTTGACCTTAGCGCCGAGACACTC CAGGCACGCCGGGAATGGGGCCCCATCTTTAATATCTGAAGGAGAAGAACTTCCAGCCACGAATCT CTTACCCTGCAAAGTTGAGTTTTATCTCCGAGGGTGAGATTAAGTATTTTCATCGATAAACAGATGCT GCGAGACTTCGTGACAACCTCGCCAGCTCTCAAGGAAGTCTCAAGAGAGGCTCTTAATATGGAGCGC AATAATAGATATCAACCTTTGAGAACACGCAAGATGgggtccggcgagggcgaggggaagCcttc taacatgcggggacgtggaggaaaatccccggcccaGGTAGCGGCccaaagaagaagcggaaggtcgg cggcggcagcACCGGCTCTAACTCACATATCACCATCCTTACACTTAACATTAACGGCTCAACTCA GCTATCAAGCGCCATCGGCTGGCCAGCTGGATCAAATCACAGGATCCAAGCGTTTGTTCATCCAAG AGACCCACCTGACCTGTAGAGATACTACCGCCTCAAGATCAAGGGATGGCGAAAGATTTATCAGGC GAACGGTAAGCAGAAGAAAGCCGGAGTCGAATTTCTGGTCTCAGACAAGACGGATTTCAAGCCCACC AAAATTAAAGCGTGATAAGGAAGGTCACTATATTATGCTGAAAGGCAGCATACAGCAGGAAGAAGTTC AATATTGAACATCTACGCGCCAAACACCGGCGACCTCGCTTTATCAAAACAGGTCTCTGTCAGATC GCAGCGAGATCTGGATTCTCATACGTTGATTATGGGTGATTTCATACACCATTGAGCAGCCCTGGAT CGCAGCACAGGCAAAAGGTAAATAAAGACACGCAAGAGCTCAATAGCGCACTGCATCAGGCAGATC TCATTGATATTTATCGCACTCTTCATCCTAAGAGTACCGAGTACACATTTCTTCAGCGCCCCACATCA

gacgttgtggctgttgttagttgtactccagcttgtgccccaggatgttgccgtcctccttgaagtgc atgcccttcagctcgatgcggttcaccaggggtgtcgccctcgaacttcacctcggcgcggttcttgt agttgccgtcgtccttgaagaagatgggtgcgctcctggacgtagccttcgggcatggcggaacttgaa gaagtcgtgctgcttcatgtgggtcggggtagcggtgaagcactgcacgccgtaggtcaggggtggtc acgaggggtggggccagggcacgggcagcttgccgggtggtgcagatgaacttcagggtcagcttgccgt aggtggcatcgccctcgccctcgccggacacgctgaacttgtggccgtttacgtcgccgtccagctc gaccaggatgggcaccaccccggtgaacagctcctcgcccttgtccaccatggtggcgggatctgac ggttcactaaaccagctctgtttatatagacctcccaccgtacacgcctaccgcccatttgcgtcaa tggggcgaggttgttacgacattttggaaagtcctcggttgcattttgggtgccccaaacaaactcccattg acgtcaatgggggtggagacttggaaatccccgtgagtcacacccgctatccacgcccattgtagtact gccccaaaccgcatcaccatggtaatagcgatgactaatacgtagatgtactgccaagtaggaaagtc ccataaggtcatgtactgggcataatgccaggcgggccatttaccgtcattgacgtcaataggggggc gtacttggcatatgatacacttgatgtactgccagtgggcagtttaccgtaaaatactccaccatt gacgtcaatggaaagtccctatttggcggttactatgggaacatacgtcattattgacgtcaatggggcg ggggtcgttggggcggtcagccaggcgggccatttaccgtaagttatgtaacgGCCTGCTGCCGGCT CTGCCGGCTCTTCCGCGTCTTCGCCCTTCGCCCTCAGACGAGTCGGATCTCCCTTTGGGCCGCCCTCCC CGCCTGTCTAGCTTGACTGACTGAGATACAGCGTACCTTCAGCTCACAGACATGATAAGATACATTG ATGAGTTTGGACAAACCACAACCTAGAATGCAGTGAAAAAATGCTTTATTTGTGAAATTTGTGATGC TATTGCTTTATTTGTAACCATTATAAGCTGCAATAAACAAGTTAACAACAACAAATTGCATTCATTTT ATGTTTCAGGTTTCAGGGGGAGGTGTGGGAGGTTTTTTTAAAGCAAGTAAAACTCTACAAATGTGGTA TTGGCCCATCTCTATCGGTATCGTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTT TTGTGCCCCCTCGGGCCGGATTGCTATCTACCGGCATTGGCGCAGAAAAAATGCCTGATGCGACGCT GCGCGTCTTATACTCCCACATATGCCAGATTGAGCAACGGATACGGCTTCCCCAACTTGCCCACTTC CATACGTGTCTCTCTTACCAGAAATTTATCCTTAAGGTGCTCAGCTATCCTGCAGGCGATCTCTCGA TTTCGATCAAGACATTCTCTTAATGGTCTTTTCTGGACACCAGTAGGGGTGAGAAGTATCATCAA ACTTTCTTCCCTCCCTAATCTCATTGGTTACCTTGGGCTATCGAAACTTAATTAAGCGATCTGCATC TCAATTAGTCAGCAACCATAGTCCCGCCCCCTAATCCGCCCCATCCCGCCCCCTAATCCGCCCCAGTTC CGCCCCATTCTCCGCCCCATCGCTGACTAATTTTTTTTTTATTTATGCAGAGGCCGAGGCCGCCCTCGGCC TCTGAGCTATTCCAGAAGTAGTGAGGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAGGAGGTAGCC AACATGATTGAACAAGATGGATTGCACGCAGGTTCTCCCGCCGCTTGGGTGGAGAGGCTATTTCGGCT ATGACTGGGCACAACAGACAATCGGCTGCTCTGATGCCGCCGTGTTCCGGCTGTCAGCGCAGGGGCG CCCGGTTCTTTTTGTCAAGACCGACCTGTCCGGTGCCCTGAATGAACTCCAGGACGAGGCAGCGCGG CTATCGTGGCTGGCCACGACGGGCGTTCCTTGCGCAGCTGTGCTCGACGTTGTCACTGAAGCGGGAA GGGACTGGCTGCTATTGGGCGAAGTGCCGGGGCAGGATCTCCTGTCATCTCACCTTGCTCTCGCGA GAAAGTATCCATCATGGCTGATGCAATGCGGCGGCTGCATACGCTTGATCCGGCTACCTGCCCATTTC GACCACCAAGCGAAACATCGCATCGAGCGAGCACGTACTCGGATGGAAGCCGGTCTTGTCGATCAGG ATGATCTGGACGAAGAGCATCAGGGGCTCGCGCCAGCCGAACCTGTTTCGCCAGGCTCAAGGCGCGGAT GCCCCAGGGCGAGGATCTCGTCTGACCCACGGCGATGCCTGCTTGCCGAATATCATGGTGGAAAAAT GGCCGCTTTTCTGGATTTCATCGACTGTGGCCGGCTGGGTGTGGCGGACCGCTATCAGGACATAGCGT TGGCTACCCGTGATATTGCTGAAGAGCTTGGCGGCGAATGGGCTGACCGCTTCTCTGCTGCTTTACGG TATCGCCGCTCCCGATTTCGACGCGCATCGCCTTCTATCGCCTTCTTGACGAGTTCTTCTAGTATGTA AGCCTGTGCTCTTAGTTGCCAGCCATCTGTTGTTTGCCCCCTCCCCCGTCCCTTCTTGACCTTG AAGGTGCCACTCCCACTGTCTTTCTTAATAAATGAGGAAATTGCATCGCATGTCTGAGTAGGTG TCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGG CATGCTGGGGATGCGGTGGGCTCTATGGTTAATTAACCAGTCAAGTCAGCTACTTGGCGAGATCGAC TTGTCTGGGTTTTCGACTACGCTCAGAATTGCGTCAGTCAAGTTCGATCTGGTCCCTTGCTATTGCACC CGTTCTCCGATTACGAGTTTCATTTAAATCATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTA AAAAGGCCGCTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCTGACGAGCATCACAAAAATCGACG CTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATAACCAGGCGTTTCCCCCTGGAAGCTCC CTCGTGCCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCTTCTCCTTCGGGAA GCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTGCTTCGCTCCAAGCT GGGCTGTGTGCACGAACCCCCGTTACGCCCCAGCGCTGCGCCTTATCCGGTAACATATCGTCTTGAG TCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGA GGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGT ATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGC AAACAAACCACCGCTGGTAGCGGTGGTTTTTTTTGTTTGAAGCAGCAGATTACGCGCAGAAAAAAG GATCTCAAGAAGATCCTTTGATCTTTTCTACGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTA AGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTACCTAGATCCTTTTAAATTAAAAATGAAGT TTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGG CACCTATCTCAGCGATCTGTCTATTTTCGTTTCATCCATAGTTGCATTTAAATTTCCGAACCTCTCAAG GCCCTCGTCGGAATACTTCAAACCTTTCGTCGATCCATCTTGCAGGCTACCTCTCGAACGAACATA TCGCAAGTCTCT (SEQ ID NO: 109)

gcacgggagcgttgccgggtggtgcagatgaacttcagggtcagcttgccgtaggtggcatcgccctc gccctcgccggacacgctgaacttgtggccgtttacgtcgccgtccagctcgaccaggatgggcacc acccgggtgaacagctcctcgcccttgctcaccatggtggcgggatctgacgggttcaactaaaccagc tctgcttatatagacctccacccgtacacgcctaccgcccatttgcgtaaatggggcgagggtgtta cgacatthttggaaagtcccgttgattttggtgccccaaactcccattgacgtcaatgggggtgga gacttggaaatccccgtgagtcacacccgtatccacgcccattgatgtactgccccaacgcacac catggtaatagcgtatgactaatacgtatgactgccaagtaggaaagtcccataagggtcatgtac tgggcataatgccaggcgggccatttaccgctcattgacgtcaatagggggcgtagcttggcataatgat acacttgatgtactgccaagtgggcagtttaccgttaaatactccaccattgacgtcaatggaaagt ccctattggcggttactatgggaacatacgtcattattgacgtcaatgggcgggggtcggtgggcggt cagccaggcgggccatttaccgtaagttatgtaacgGGCCTGCTGCCGGCTCTGCCGGCTCTTCCGC GTCTTCGCCTTCGCCCTCAGACGAGTCGGATCTCCCTTTGGGCGCCTCCCCGCTGTCTAGCTTGA CTGACTGAGATACAGCGTACCTTCAGCTCACAGACATGATAAGATACATTGATGAGTTTGGACAAAC CACAAGTAGAATGCAGTGAAAAAATGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTATTTGTA ACCATTATAAGCTGCAATAAACAAGTTAACAACAACAATTGCATTCAATTTATGTTTCAGGTTTCAGG GGGAGGTGTGGGAGGTTTTTTTAAAGCAAGTAAAACCTCTACAAATGTGGTATTGGCCATCTCTATC GGTATCGTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGTGCCCTCGGGCC GGATTGCTATCTACCGGCATTGGCGCAGAAAAAATGCCTGATGCGACGCTGCGCGTCTTATACTCC CACATATGCCAGATTGAGCAACGGATACGGCTTCCCCAAGTTGCCCACTTCCATACGTGTCTCCTT ACCAGAAATTTATCCTTAAGGTGCTGAGCTATCCTGCAGGCGATCTCTCGATTTCGATCAAGACATT CCTTTAATGGTCTTTTCTGGACACCACTAGGGGTGAGAAGTAGTTTCATCAAACTTTCTTCCCTCCCT AATCTCATTGGTTACCTTGGGCTATCGAACTTAATTAAGCGATCTGCATCTCAATTAGTCAGCAAC CATAGTCCCGCCCCCTAACTCCGCCCATCCCGCCCCCTAACTCCGCCCAGTTCCGCCCATTCTCCGCC CATCGCTGACTAATTTTTTTTTTATTTATGCAGAGGCGGAGGCGCCTCGGCCCTGTGAGCTATTCCAGA AGTAGTGAGGAGGCTTTTTTGGAGGCTTAGGCTTTTGCAAAGGAGGTAGCCAACATGATGAACAAG ATGGATTGCACGCAGGTTCTCCCGCCGCTTGGGTGAGAGGCTATTTCGGCTATTGACTGGGCACAACA GACAATCGGCTGCTCTGATGCCGCCGTGTTCGGCTGTGACGCGCAGGGGCGCCGGTTCTTTTTGTCT AAGACCGACCTGTCCGGTGCCCTGAATGAAGTCCAGGACGAGGCAGCGCGGCTATCGTGCGTGGCCA CGACGGGCGTTCTTGGCGAGCTGTGCTCGACGTTGTCACTGAAGCGGGAAGGGACTGGCTGCTATT GGGCGAAGTGCCGGGGCAGGATCTCCTGTCTCATCTCACCTTGCTCCTGCCGAGAAAGTATCCATCATG GCTGATGCAATGCGGCGGCTGCATACGCTTGATCCGGCTACCTGCCCATTCGACCACCAAGCGAAAC ATCGCATCGAGCGAGCACGTACTCGGATGGAAGCCGGTCTTGTGATCAGGATGATCTGGACGAAGA GCATCAGGGGCTCGCGCCAGCCGAAGTGTTCGCCAGGCTCAAGGCGCGGATGCCGACGGCGAGGAT CTCGTCGTGACCCACGGCGATGCCTGCTTGCCGAATATCATGGTGGAATATGGCCGCTTTTCTGGAT TCATCGACTGTGGCCGGCTGGGTGTGGCGGACCGCTATCAGGACATAGCGTTGGCTACCCGTGATAT TGCTGAAGAGCTTGGCGGCGAATGGGCTGACCGCTTCCTCGTGCTTTACGGTATCGCCGCTCCCGAT TCGACGCGCATCGCCTTCTATCGCCTTCTTGACGAGTTCTTCTAGTATGTAAGCCCTGTGCCTTCTA GTTGCCAGCCATCTGTTGTTTGCCCCCTCCCCCGTGCCTTCTTACCCCTGGAAGGTGCCACTCCAC TGTCCTTTTCTAATAAAATGAGGAAATGCATCGCATTGTCTGAGTAGGTGTCTATTCTTGGGG GGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGG TGGGCTCTATGGTTAATTAACCAAGTCAAGTCAGTACTTGGCGAGATCGACTTGTCTGGGTTTCGAC TACGCTCAGAATTGCGTCAGTCAAGTTCGATCTGGTCTTGTCTATTGCACCCGTTCTCCGATTACGA GTTTCATTAAATCATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCTTGCT GGCGTTTTTTCATAGGCTCCGCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGG CGAAACCCGACAGGACTATAAAGATACAGGCGTTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTG TTCCGACCCTGCCGCTTACCGGATACCTGTCCGCTTTCTCCCTTCGGGAAGCGTGCGCTTTCTCA TAGCTCACGCTGTAGGTATCTCAGTTCCGGTGTAGGTGCTTCGCTCCAAGCTGGGCTGTGTGCACGAA CCCCCGTTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGAC ACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGC TACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCT CTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTG GTAGCGGTGGTTTTTTTTTGTGTTGCAAGCAGCAGATTACGCGCAGAAAAAAGGATCTCAAGAAGATCC TTTGATCTTTTCTACGGGTCTGACGCTCAGTGGAAACGAAAACTCACGTTAAGGGATTTTGGTCATG AGATTATCAAAAAGGATCTTACCTAGATCCTTTTAAATTAATAAATGAAGTTTTTAAATCAATCTAAA GTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGAT CTGTCTATTTCTGTTTCATCCATAGTTGCATTTAAATTTCCGAAGTCTCCAAGGCCCTCGTCGGAAAAAT CTTCAAACCTTTCTGTCGATCCATCTTGCAGGCTACCTCTCGAACGAAGTATCGCAAGTCTCTTGGC CGGCCTTGGCCTTGGCTATTGCTTGGCAGCGCCTATCGCCAGGTATTACTCCAATCCCGAATATCC GAGATCGGGATCACCCGAGAGAAGTTCAACCTACATCCTCAATCCCGATCTATCCGAGATCCGAGGA ATATCGAAATCGGGGCGCGCCTGGTGT (SEQ ID NO: 110)
--