

METHODS AND COMPOSITIONS FOR GENOMIC INTEGRATION

CROSS REFERENCE

[0001] This Application claims the benefit of U.S. Provisional Application No. 62/895,441, filed on September 3, 2019, U.S. Provisional Application No. 62/908,800, filed on October 1, 2019, and U.S. Provisional Application No. 63/039,261, filed on June 15, 2020, each of which is incorporated herein 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

[0002a] In one aspect, provided herein is a composition comprising a human cell comprising modified genomic DNA, wherein the modified genomic DNA comprises a human insert sequence, wherein the human insert sequence comprises: (a) a DNA sequence encoding an exogenous human therapeutic polypeptide, and (b) a DNA sequence comprising a reverse complement of a human mobile genetic element sequence, wherein the human mobile genetic element sequence is an RNA sequence of the RNA molecule comprising a sequence encoding a polypeptide with target-primed reverse transcription (TPRT) activity, and wherein the human cell expresses the exogenous human therapeutic polypeptide from the human insert sequence.

[0002b] In another aspect, provided herein is a method of expressing an exogenous therapeutic polypeptide

from a genomically integrated DNA sequence of a target human cell population, the method comprising: (a) contacting an RNA molecule to the target human cell population, wherein the RNA molecule comprises: (i) an RNA sequence that is a reverse complement of the DNA sequence encoding the exogenous human therapeutic polypeptide, and (ii) a human mobile genetic element comprising an RNA sequence encoding a polypeptide with target-primed reverse transcription (TPRT) activity; and (b) translating the human mobile genetic element to produce the polypeptide with TPRT activity; (c) reverse transcribing the RNA sequence that is the reverse complement of the DNA sequence encoding the exogenous human therapeutic polypeptide via the TPRT activity of the polypeptide translated in step (b), thereby producing the DNA sequence encoding the exogenous human therapeutic polypeptide; (d) integrating the DNA sequence encoding the exogenous human therapeutic polypeptide into the genomic DNA of a target human cell of the target human cell population at a target site; and (e) expressing the exogenous human therapeutic polypeptide from the genomically integrated DNA sequence encoding the exogenous human therapeutic polypeptide in the target human cell.

[0002c] In another aspect, provided herein is a composition comprising an isolated or purified RNA molecule for use as a medicament, wherein the RNA comprises (i) an RNA sequence that is a reverse complement of the DNA sequence encoding the exogenous human therapeutic polypeptide, and (ii) a human mobile genetic element comprising an RNA sequence encoding a polypeptide with target-primed reverse transcription (TPRT) activity.

[0002d] In another aspect, provided herein is use of a composition comprising an isolated or purified RNA molecule in the manufacture of a medicament for treating a disease or condition in a human subject, wherein the RNA molecule comprises (i) an RNA sequence that is a reverse complement of a DNA sequence encoding an exogenous human therapeutic polypeptide; and (ii) a human mobile genetic element comprising an RNA sequence encoding a polypeptide with target-primed reverse transcription (TPRT) activity.

[0003] Provided herein are compositions and methods for stable, non-viral transfer and integration of genetic material into a cell. In one aspect, the genetic material is a self-integrating polynucleotide. The genetic material can be stably integrated in the genome of the cell. The cell may be a human cell. The method is designed for a safe and reliable integration of a genetic material into the genome of a cell.

[0004] In one aspect, provided herein are compositions and methods that allow integration of genetic material into the genome of a cell, wherein the genetic material that can be integrated is not specifically restricted by size. In some aspects, the method described herein provides a one-step, single polynucleotide-mediated delivery and integration of genetic "cargo" in the genome of a cell. The genetic material may comprise a coding sequence, e.g., a sequence encoding a transgene, a peptide, a recombinant protein, or an antibody or fragments thereof, wherein the method and compositions ensure stable expression of the transcribed product encoded by the coding sequence. The genetic material may comprise a non-coding sequence, for example, a regulatory RNA sequences, e.g., a regulatory small inhibitory RNA (siRNA), microRNA (miRNA), long non-coding RNA (lncRNA), or one or more transcription regulators such as a

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).

[0051] **FIG. 1B** 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.

[0052] **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.

[0053] **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.

[0054] **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.

[0055] **FIG. 3A** illustrates an exemplary diagram of conventional circRNA structure and formation.

[0056] **FIG. 3B** illustrates two views of an exemplary RL-GAAA tectoRNA motif designs (SEQ ID NOS 80-81, respectively, in order of appearance).

[0057] **FIG. 3C** illustrates exemplary structures of chip-flow piece RNAs as platforms for testing potential tectoRNA.

[0058] **FIG. 4A** illustrates an exemplary schematic showing ORF2p binding to an ORF2 poly A region.

[0059] **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.

[0060] **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.

[0061] **FIG. 5** illustrates exemplary constructs for integrating an mRNA encoding a transgene into the genome of a cell.

[0062] **FIG. 6A** illustrates an exemplary construct with a sequence encoding ORF1p for integrating an mRNA encoding a transgene into the genome of a cell.

[0063] **FIG. 6B** illustrates an exemplary construct without a sequence encoding ORF1p for integrating an mRNA encoding a transgene into the genome of a cell.

[0064] **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 NO: 82) in the 5'UTR; and/or xrRNAs, a triplex motifs (SEQ ID NOS 83-85, respectively, in order of appearance) and/or a non-A nucleotide residues in the 3'UTR.

[0065] **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.

[0066] **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.

[0067] **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.

[0068] **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.

[0069] **FIG. 8C** shows exemplary flow cytometry results from the results shown in **FIG. 8B**.

[0070] **FIG. 9A** 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 a recombinant chimeric fusion receptor protein (ATAK receptor) that has extracellular region capable of binding to CD5 and an intracellular region comprising an FCR intracellular domain and a PI3 kinase recruitment domain. The coding sequence of the ATAK receptor is interrupted with an intron.

[0071] **FIG. 9B** shows exemplary results showing successful integration of the mRNA transcript encoded by the plasmid shown in **FIG. 9A** and expression of ATAK relative to mock-transfected cells (fold increase in mean fluorescence intensity of ATAK positive cells is shown). Mock transfected cells were transfected by the vector lacking the ATAK cargo sequence. Expression of ATAK receptor protein was detected by binding with a labeled CD5 antibody.

[0072] **FIG. 9C** shows exemplary flow cytometry results from the results shown in **FIG. 9B**.

[0073] **FIG. 10A** 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 a recombinant chimeric fusion receptor protein (ATAK receptor) followed by a T2A self-cleavage sequence followed by a split GFP sequence (all in a reverse orientation relative to the LINE-1 sequence). The coding sequence

hyperactive Tn5 transposase that is active in an *in vitro* transposition reaction comprises a transferred strand that exhibits a transferred transposon end sequence as follows: 5' AGATGTGTATAAGAGACAG 3' (SEQ ID NO: 51), and a non-transferred strand that exhibits a "non-transferred transposon end sequence" as follows: (5' CTGTCTCTTATACACATCT 3' (SEQ ID NO: 52). 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.

[00141] 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

[00142] 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).

[00143] 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

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 MGKKQNRKTGNSKTQSASPPPKERSSSPATEQSWMENDFDELREEGFRRSNYSELREDIQTGKKEVENFEKNLEECITRITNTEKCLKELMELKTKARELREECRSLRSRCDQLEERVVSAMEDEMNMKREGKFREKRIKRNEQSLQEIWDYVKRPNLRLIGVPESDVENGTKLENTLQDIIQENFPNLARQANVQIQEIQRTPQRYSSRRATPRHIIVRFTKVEMKEKMLRAAREKGRVTLKGKPIRLTVDLAETLQARREWGPIFNILKEKNFQPRISYPAKLSFISEGEIKYFIDKQMLRDFVTTRPALKELLKEALNME RNNRYQPLQNHAKM (SEQ ID NO: 53). 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 atgggcaagaagcaaaatcgcaagacggggaattccaagacacaatccgtagccaccacctaagagcgttctagctcccctgctactgagcagctctggatggaaaacgacttcgatgaactccgggaagagggaattagcgcatccaactattcagaactccgcgaagatatccagacaaaggggaaggaa gtcgagaatttcgagaagaacctcgaggagtgcatcacccgtatcacaacactgagaaatgtctcaagaactcatggaacttaagacaaaagccagg gagcttcgagaggagtgctggagtctgagatccaggtgtgaccagctcgaggagcgcgtgagcgcgatggaagacgagatgaacgagatgaaaag agagggcaaatcaggagagaagcgcattaagaggaacgaacagagtctgcaggagatttgggattacgtcaagaggcctaactgcggtgatcgcc gtccccgagagcgacgtagaaaacgggactaaactggagaatacacttcaagacatcattcaagaaaattttccaaacctggctcggaagctaattgtg caaatccaagagatccaacgcacacccagcgggtatagctctcggcgtgccaccctaggtatattatctgctgctttactaagggtggagatgaaagag aagatgctgcgagccgctcgggaaaagggaagggtgactttgaagggcaaacctattcgggtgacggttgacctagcgcgagacactccaggcac gccgggaatggggcccatctttaatatcctgaaggagaagaactccagccacgaatctctaccctgcaaagttgagtttatctccgaggggtgagatta agtatttcatcgataaacagatgctgcgagacttcgtgacaactcgcggcgtctcaaggaactgctcaagaggctcttaatatggagcgcaataataga tatcaacccttcgagaaccacgcaaagatgtga (SEQ ID NO: 54).

[00217] 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 MTGSNSHITILTLNINGLNSAIKRHRLASWIKSQDPSVCCIQETHLTCRDTHRLKIKGWRKIYQAN GKQKKAGVAILVSDKTDFKPTKIKRDKEGHYIMVKGSIQQEELTILNIYAPNTGAPRFIKQVLSDL QRDLDSTLIMGDFNTPLSTLDRSTRQKVNKDTQELNSALHQADLIDIYRTLHPKSTEYTFFSAP HHTYSKIDHIVGSKALLSKCKRTEIITNYLSDHSAIKLELRIKNLTQSRSTTWKLNNLLNNDYWV HNEMKAEIKMFFETNENKDDTTYQNLWDAFKAVCRGKFIALNAYKRKQERSKIDTLTSQLKELE KQEQTTHSKASRRQEITKIRAELEKETQKTLQKINESRSWFFERINKIDRPLARLIKKKREKNQIDTI KNDKGDITTDPTIEIQTIREYYKHL YANKLENLEEMDTFLDTYTLPRLNQEEVESLNRPTGSEIV AIINSLPTTKKSPGPDGFTAIFYQRYMEELVPFLKLFQSIEKEGILPNSFYEASIIIPKPGRDTTKKE NFRPISLMNIDAKILNKILANRIQQHIKKLIHHDQVGFIPGMQGWFNIRKSINVIQHINRAKDKNH

MIISIDAEKAFDKIQPFMLKTLNKLIDGTYFKIIRAIYDKPTANIILNGQKLEAFPLKTGTRQGCPLSP
 LLFNIVLEVLARAIQEKEIKGIQLGKEEVKLSLFADDMIVYLENPIVSAQNLLKLISNFSKVSQYKINV
 QKSQAFLYTNNRQTESQIMGELPFVIAASKRIKYLGIQLTRDVKDLFKENYKPLLKEIKEDTNKWKNIP
 CSWVGRINIVKMAILPKVIYRFNAIPIKLPMTFFTELEKTTLKFIWNQKRARIAKSILSQKNKAGGITLP
 DFKLYYKATVTKTAWYWYQNRDIDQWNRTEPSEIMPHIYNYLIFDKPEKNKQWGKDSL FNKWCWE
 NWLAICRKLKLDPFLTPYTKINSRWIKDLNVKPKTIKTLEENLGITIQDIGVGKDFMSKTPKAMATKD
 KIDKWDLIKLSFCTAKETTIRVNRQPTTWEKIFATYSSDKGLISRIYNELKQIYKKKTNNPIKKWAKD
 MNRHFSKEDIYAAKKHMKKCSSSLAIREMQIKTTMRYHLTPVRMAIKKSGNNRCWRGCGEIGTLLH
 CWWDCKL VQPLWKS VWRFLRDLELEIPFDPAILLGIYPNEYKSCCYKDTCTRMFIAALFTIAKTWN
 QPKCPTMIDWIKKMWHIYTMEYYAAIKNDEFISFVGTWMKLETIILSKLSQEQTCKHRIFSLIGGN

(SEQ ID NO: 55). 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
 atgaccggctctaactcacatataccatccttacacttaacattaacggcctcaactcagctatcaagcgccatcggtggccagctggatcaaatcacaggatc
 caagcgtttgtgcatccaagagaccacctgacgttagagatactaccgctcaagatcaaggatggcgaaagatttatcaggcgaacggtaagcagaa
 gaaagccggagtcgcaattctggtctcagacaagacggattcaagcccacaaaattaagcgtgataaggaagtcactatattatggtgaaaggcagcatatc
 agcaggaagaacttaccatattgaactctacgcgcaaacacccggcgacctcgtttatcaaacaggtcctgtccgagctgcagcgagatctggtattctata
 cgttgattatgggtgattcaatacaccattgagcacctggatcgagcaccaggcaaaaggtaaataaacacgcaagagctcaatagcgactgcactgcacg
 gcagatctcattgatatttatcgactcttcatcctaagagtaccgagtacacattctcagcgcccatcatatactcaaatcgcacatctgcggctca
 aggtctgctgtaaaagtgaagcgacagagataattacaaattacctgtcagatcatagcgcatgaagctcgagctgagaatcaagaacctgaccagag
 ccggagtaccacttggaagcttaataacctgctgctcaacgattattgggtccacaatgagatgaaggcagagattaaaatgttttcgaaacaaatgagaataag
 gatactacatcaaaaccttgggatgccttaagccgctgctgagaggcaagttcatcgccctcaacgcctataaaagaaaacaagagagatctaagatcgat
 actctcacctctcagctgaaggagttggagaaacaggaacagaccactccaaggcgtaagacggcaggagatcacaagattcgccggagttgaaga
 gatcgaaacccaaaagactcttcagaaaattaacgagtcctgtagttggtttcgcagcgattaataagatagacagacctctggcacgactgattaagaagaa
 gcgcgaaaagaaccagattgatacatcaagaacgacaaggcgacatcactactgacccgaccgagatccagaccactattcgggagttattataagcattgt
 atgctaacaagcttgagaacctggaagagatggacacttttctgatactatactctgccacggcttaataagagggaagtcgagtcctcaaccgccaattac
 aggaagcgagattgtggccataaataacctctgccgacaagaaatctcctggtccggacgggtttacagctgagttttatcaacggtatatggaagagcttgta
 ccgtttctgctcaagctcttcagctctatagaaaaggaaggcatcttgcccaattcctctacgaagctctataatacttattcccaaaccaggacgcgataaccaca
 agaaggaaaacttccggccattagctcatgaatatcgacgctaaaatattgaacaagattctcgccaacagaatccaacaacatattaagaaattgatacatca
 cgaccaggtggggtttatactggcatgcagggtggttaacatccggaagagtattaacgtcattcaacacattaatagagctaaggataagaatcatatgatc
 atctctatagacgcggaaaaggcattcgataagattcagcagccatttatgctcaagactctgaacaaactcggcacgcggaacatatttaagattattcgcgc
 aatttacgataagccgactgctaacattatccttaacggccaaaagctcgaggccttccgctcaagactggaacccgccaaggctgtcccctctccccgctttgt
 ttaatatgtactcgaggtgctggctagggtattctgaagagaaaagattaaaggatacagctcggaagggaagaggtcaagcttccctgttcgccgatg
 atatgattg

tgtacctggagaatcctattgtgtctgctcagaaccttcttaacttatttctaacttttagcaaggtcagcggctataagattaacgtccagaaatctcaggcctttctgt
acacaaataatcgacagaccgaatcccagataatgggtgagcttccgtttgcatagccagcaaaaggataaagtatctcggaatccagctgacacgagacgtt
aaagatttgtttaagggaaaattacaagcctctcctgaaagagattaaggaagatactaataagtggagaatatccctgttcatgggtggcagaatcaacatagt
gaagatggcaatacttctaaagtgatatacgctttaacgccatccaattaactgcctatgaccttcttacggagctcgagaaaacaaccttaaatatatgg
aatcaaaagagagcaagaatagcgaagtccatcttgagccagaagaataaggccgggtgggattactttgcctgattttaagttgtattataaagccacagtaacta
agacagcctgggtattgggtatcagaatagacatcgaccagtgggaatcgaccgaacctatcagagataatgccccacatctataattaccttatattcgataagcc
agaaaagaataaacagtggggcgaagacagcctcttcaacaagtgggtgtgggagaattggctggccatatgccggaaactcaagctcgaccttttcttacac
cctacactaaaataacagtaggtggatcaaggacttgaatgtcaagccaaagactataaagacactggaagagaatctgggatcacaatacagaatataggc
gtcggcaagattttatgtcaaagacgcccgaaggccatggccactaaggataagattgataagtgggaccttattaagctcaaaagcttctgtactgccaaggag
accacgatcagagttaataggcagcccactacatgggaaaagatttccgacctattcatcagataaggggtgataagcagaatatataacgagctgaagcag
atctacaagaagaaaacgaataatcccatcaagaagtgggcaaaagatatgaacaggcattttagcaagaggatatctacgccgcaagaagcatatgaaga
agtgtagttaagcttggccattcgtgagatgcagattaagacgacctatgccataaccctacccagtgaggatggcaattatcaagaaatctggcaataatag
atgttggcggggctgtggcgagattggcaccctgctccattgctgggtgggattgcaagctgggtgcagccgtttggaatcagctctggcgctttctgagggacct
cgagcttgagattcccttcgatcccgaattcccttgcctggaaatctatcctaagcaatacagaagctgttgttacaagatacgtgtacccggatgttcacgcgg
ccttgtttacgatagctaagacgtggaatcagcctaagtccccacaatgatcgattggatcaagaaaatgtggcatattatcatcaggagattacgcagcaatt
aagaatgacgaatttatttctcgttgggacctggatgaagctggagactattattctgagcaagctgtctcaggagcaaaagacaaaagcatagaatcttctctc
attgggtgtaactaa (SEQ ID NO: 56).

[00218] 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 MVIGTYISITLNVNGLNAPTKRHRLAEWIQKQDPYICCLQETHFRPRDTYRLKVRGWKKIFHANGNQ KKAGVAILISDKIDFKIKNVTRDKEGHYIMIQGSIQEEDITIINIYAPNIGAPQYIRQLLTAIKEEIDSNTII VGDfNTSLTPMDRSSKMINKETeALNDTIDQIDLIDIYRTFHPKTADYTFSSAHGTFsRIDHILGHKS SLSKFKKIEIISsIFSDHNAMRLEMNHREKNVKKTNtwRLNNTLLNNQEITEEIKQEIKKYLETNDNEN TTTQNLWDAAKAVLRGKFIAIQAYLKKQESQVNNLTLHLKKLEKEEQTKPKVSRRKEIKIRAEINE IETKKTIAKINKTKSWFFeKINKIDKPLARLIKkKRERTQINKIRNEKGEVTTDTAEIQNILRDYYKQLY ANKMDNLEEMDKFLERYNLPRLNQEETENINRPITSNEIETVIKNLPTNKSPGPDGFTGEFYQTFREEL TPILLKLFQKIAEEGTLpNSFYeATITLIPKPKDttKKENYRPISLMNIDAKILNKILANRIQQHIKRIIH HDQVGFIpGMQGFfNIRKSINVIHHINKLKKKNHMIISIDAEKAFDKIQHPfMIKTLQKVGIEGTYLNII KAIYDKPTANIILNGEKLKAFPLRSgTRQGCPLSPLLFNIVLEVLATAIREEKEIKGIQIGKEEVKLSLFA DDMILYIENPKTATRKLLELINEYgKVAGYKINAQKSLAFLYTNDeksEREIMETLPFTIATKRIKYLGI NLPKETKDLYAENYKTLmKEIKDDTNRWRDIPCSWIGRINIVKMSILPKAIYRFNAIPIKLPMaFFTEL EQIILKFVWRHKRPRIAKAVLRQKNGAGGIRLPDFRLYYKATVIKTIWYWHKNRNIDQWNKIESPEIN PRtYGQLIYDKGGKDIQWRKDSLF

NKWCWENWTATCKRMKLEYSLTPYTKINSKWIRDLNIRLDTIKLLEENIGRTLFDINHSKIIFDPP
 PRVMEIKTKINKWDLMKLQSFCTAKETINKTKRQPSEWEKIFANESTDKGLISKIYKQLIQLNIKE
 TNTPIQKWAEDLNRHFSKEDIQTATKHMKRCSTSLIIREMQIKTTMRYHLTPVRMGIRKSTNNK
 CWRGCGEKGTLHLCWWECKLIQPLWRTIWRFLKKLKIELPYDPAIPLLGITYPEKTVIQKDTCTR
 MFIAALFTIARSWKQPKCPSTDEWIKKMWYIYTMEYSSAIKRNEIGSFLETWMDLETVIQSEVSQ
 KEKNKYRILTHICGTWKNGTDEPVC RTEIETQM (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

atggcataggaacatacatatcgataattacctaactgaatggattaaatgccccacaaaagacatagactggctgaatggatacaaaacaaga
 cccatataatgctgtctacaagagaccacttcagacctaggacacatacagactgaaagtgaggggatggaaaaagattccatgcaaatggaaat
 caaaagaaagctggagtagctatactcatatcagataaaatagactttaaataaagaatgttacaagagacaaggaaggacactacataatgatccagg
 gatcaatccaagaagaagatataacaattataatataatgacccaacataggagcacctcaatacataaggcaactgctaacagctataaaagagga
 aatcgacagtaacacaataatgtgggggactttaacacctcacttacaccaatggacagatcatcaaaatgaaaataaataaggaaacagaagcttta
 aatgacacaatagaccagatagatttaattgatataatagacattccatcaaaaacagcagattacacgttcttcaagtgcgcacggaacattctcca
 ggatagatcacatcttgggtcacaaatcaagcctcagtaaatttaagaaaattgaatcatatcaagcatctttctgaccacaacgctatgagattagaat
 gaatcacagggaacacgtaaaaaagacaaacacatggaggctaaacaatacgtfactaaataaccaagagatcactgaagaaatcaaacaggaa
 ataaaaaatacctagagacaaatgacaatgaaaacacgacgacccaaaacctatgggatgcagcaaaagcgggtctaagagggaagtttatagctata
 caagcctacctaagaaacaagaaaaatctcaagtaacaatctaacttacacctaagaaactagagaaagaagaacaaacaaacccaaagttag
 cagaaggaaagaaatcataaagatcagagcagaataaatgaaatagaacaaagaaacaatagcaaaatcaataaaactaaaagttgttcttga
 gaagataacaaaattgataagccattagccagactcatcaagaaaaagaggagaggactcaaatcaataaaatcagaatgaaaaaggagaagtta
 caacagacaccgcagaatacaaaacatcctaagagactactacaagcaactttatgccataaaatggacaacctggaagaaatggacaattcttag
 aaaggataacctccaagactgaaccaggaagaaacagaaaatatcaacagaccaatcacaagtaatgaaattgaaactgtgattaaaaatctccaac
 aaacaaaagtccaggaccagatggcttcacaggtgaattctatcaaacatttagagaagactaacacccatcttctcaaaactctccaaaaatgacg
 aagaaggaaactcccaaaactattctatgaggccaccatcacctgataccaaaaccagacaaagacactacaaaaaagaaaattacagaccaatat
 cactgatgaatatagatgcaaaaatcctcaacaaatactagcaaacagaatccaacaacacattaaaaggatcataccacgatcaagtgggatttatc
 ccagggatgcaaggattcttcaatatacgaatcaatcaatgtgataccatattacaattgaagaagaaaaaccatgatcatctcaatagatgca
 gaaaaagctttgacaaaattcaacacccatttatgataaaaactctccagaaagtgggcatagagggaacctacctcaacataataaaggccatatatga
 caaacccacagcaaacatcttctcaatggtgaaaaactgaaagcatttctctaatgacggaacgagacaaggatgtccactctcaccactattattca
 acatagttctggaagtcttagccacggcaatcagagaagaaaaagaataaaaggaaatacaaatggaaaagaagaagtaaaactgtcactgtttgcgg
 atgacatgatactatacatagagaatcctaaaactgccaccagaaaactgctagagctaattaatgaatatggtaaagttgcaggttacaaaatgaacac
 agaaatctcttgattcctatacactaatgatgaaaaatctgaaagagaaattatggaaacactcccatttaccattgcaacaaaaagaataaaatcctagg
 aataaacctacctaaggagacaaaagacctgtatgcagaaaactataagacactgatgaaagaaattaaagatgataccaacagatggagagatatacc
 atgttcttgattggaagaatcaacattgtgaaatgagtatactaccaaaagcaatctacagattcaatccctatcaaataccaatggcatttttac

ggagctagaacaaatcatcttaaaattgtatggagacacaaaagaccccgaaatagccaaagcagcttgaggcaaaaaatggagctggaggaatcagactccctgacttcagactatactacaagctacagtaatcaagacaatatggtactggcacaaaaacagaacatagatcaatggaacaagatagaaagcccagagattaaccacgcacctatggtcaactaatctatgacaaaggaggcaagatatacaatggagaaaagacagctcttcaataagtgtgctgggaaaactggacagccacatgtaaaagaatgaaattagaatactccctaaccatacacaaaaataaactcaaatggattagagacctaaatataagactggacactataaaactcttagaggaaaacataggagaacactcttgacataaatcacagcaagatcttttcgatccacctcctagagtaatggaaataa aaacaaaaataaacaagtgggacctaataaacttcaaagctttgcacagcaaaggaaaccataaacaagacgaaaagacaaccctcagaatgggagaaaatatttgcaaatgaatcaacggacaaaggattaatctccaaatatataaacagctcattcagctcaatatcaaaagaaacaaacacccaatccaaa atgggcagaagacctaataagacatttctccaaagaagacatacagacggccacgaagcacatgaaaagatgctcaacatcactaattattagagaaat gcaaatcaaaactacaatgaggtatcacctcactctgttagaatgggcatcatcagaaaatctacaacaacaatgctggagagggtgtggagaaaa gggaacctcttgactgttggtgggaatgtaattgatacagccactatggagaacaatatggaggtccttaaaaaactaaaaatagaattaccatatga ccagcaatcccactactgggcataatcccagagaaaaccgtaattcaaaaagacacatgcacccgaatgttcattgcagcactattacaatagccagg tcatggaagcaacctaaatgcccatcgacagacgaatggataaagaagatgtggtacatatatacaatggaatattactagccataaaaaggaacgaa ttgggtcatttttagagacgtggatggatctagagactgtcatcacagagtgaagtaagtcagaaagagaaaaacaatatcgtatattaacgcatatatgtg gaacctggaaaaatggtacagatgaaccggtctgcaggacagaaattgagacacaaatgtaa (SEQ ID NO: 58).

[00219] 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: 59). 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: 60). 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.

[00220] 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 MELTIPIFATNSDCELVKAMQGLLKDGNPIPSAIAANS GIYAMASNFTQFVLVD NGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEV PKGAWRSYLN MELTIPIFATNSDCELVKAMQGLLKDGNPIPSAIAANS GIY (SEQ ID NO: 61). In some embodiments, the MS2 coat protein sequence is fused to the ORF2p sequence.

[00221] In some embodiments, the transgene may comprise a flanking sequence which comprises an Alu ORF2p recognition sequence.

[00248] In some embodiments, the mutation is a deletion. In some embodiments, the deletion is complete, i.e., 100% of the L1 endonuclease domain is deleted. In some embodiments, the deletion is partial. In some embodiments, the about 98%, about 95%, about 94%, about 93%, about 92% about 91%, about 90%, about 85%, about 80%, about 75%, about 70%, about 65%, about 60%, or about 50% of the ORF2 endonuclease domain is deleted.

[00249] 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.

[00250] 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.

[00251] In some embodiments, the additional endonuclease domain from a heterologous protein is an endonuclease domain from a related retrotransposon.

[00252] 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 ("LAGLIDADG" disclosed as SEQ ID NO: 62) (LHEs) 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.

[00253] 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.

[00254] In some embodiments, the engineered or modified L1-ORF2p targets a recognition site that is other than the usual (Py)_n↓(Pu)_n site. In some embodiments, engineered L1-ORF2p targets a recognition site that comprises the (Py)_n↓(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↓(Pu)_n site, for example TTTT/AAG, or TTTT/AAC, or TTTT/AAT, TTTT/AAA, GTTTT/AA, CTTT/AA, ATTTT/AA, or TTTTT/AA. In some embodiments, the

comprises an scFv comprising (i) a variable heavy chain (VH) sequence with at least 90% sequence identity to EIQLVQSGGGLVKPGGSVRISCAASGYTFTNYGMNWVRQAPGKGLEWMGWINTHTGEPTYADSFK GRFTFSLDDSKNTAYLQINSLRAEDTAVYFCTRRGYDWYFDVWGQGTTVTV (SEQ ID NO: 63); and (ii) a variable light chain (VL) sequence with at least 90% sequence identity to DIQMTQSPSSLSASVGDRVTITCRASQDINSYLSWFQQKPGKAPKTLIYRANRLESGVPSRFSGSGSGT DYTLTISLQYEDFGIYYCQQYDESPWTFGGGKLEIK (SEQ ID NO: 64). 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: 65). 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 YEDMRGILYAAPQLRSIRGQPGPNHEEDADSYENM (SEQ ID NO: 66). 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 KVAKKPTNKAPHPKQEPQEINFPDDLPGSNTAAPVQETLHGCQPVQTQEDGKESRISVQERQ (SEQ ID NO: 67). 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 IYIWAPLAGTCGVLLLSLVIT (SEQ ID NO: 68). 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: 69). 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 GSGGGGSDIQMTQSPSSLSASVGDRVITICRASQDINSYLSWFQQKPGKAPKTLIYRANRLESG VPSRFSGSGSGTDYTLTISSLQYEDFGIYYCQQYDESPWTFGGGTKLEIKSGGGGSGALSNSIMYF SHFVPVFLPAKPTTTPAPRPPTPAPTIASQPLSLRPEACRPAAGGAVHTRGLDIYIWAPLAGTCGV LLLSLVITLYCRRLKIQVRKAAITSYEKSDGVYTGLSTRNQETYETLKHEKPPQSGSYEDMRGI LYAAPQLRSIRGQPGPNHEEDADSYENM (SEQ ID NO: 70). In some embodiments, the fusion protein comprises a sequence with at least 90% sequence identity to EIQLVQSGGGLVKPGGSVRISCAASGYTFTNYGMNWVRQAPGKGLEWMGWINTHTGEPTYAD SFKGRFTFSLDDSKNTAYLQINSLRAEDTAVYFCTRRGYDWYFDVWGQGTTVTVSSGGGGSGG GSGGGGSDIQMTQSPSSLSASVGDRVITICRASQDINSYLSWFQQKPGKAPKTLIYRANRLESG VPSRFSGSGSGTDYTLTISSLQYEDFGIYYCQQYDESPWTFGGGTKLEIKSGGGGSGALSNSIMYF SHFVPVFLPAKPTTTPAPRPPTPAPTIASQPLSLRPEACRPAAGGAVHTRGLDIYIWAPLAGTCGV LLLSLVITLYCRLKIQVRKAAITSYEKSDGVYTGLSTRNQETYETLKHEKPPQKKVAKKPTNKAP HPKQEPQEINFPDDLPGSNTAAPVQETLHGCQPVTQEDGKESRISVQERQ (SEQ ID NO: 71) or EIQLVQSGGGLVKPGGSVRISCAASGYTFTNYGMNWVRQAPGKGLEWMGWINTHTGEPTYAD SFKGRFTFSLDDSKNTAYLQINSLRAEDTAVYFCTRRGYDWYFDVWGQGTTVTVSSGGGGSGG GSGGGGSDIQMTQSPSSLSASVGDRVITICRASQDINSYLSWFQQKPGKAPKTLIYRANRLESG VPSRFSGSGSGTDYTLTISSLQYEDFGIYYCQQYDESPWTFGGGTKLEIKSGGGGSGALSNSIMYF SHFVPVFLPAKPTTTPAPRPPTPAPTIASQPLSLRPEACRPAAGGAVHTRGLDIYIWAPLAGTCGV LLLSLVITLYCRRLKIQVRKAAITSYEKSDGVYTGLSTRNQETYETLKHEKPPQKKVAKKPTNKA PHPKQEPQEINFPDDLPGSNTAAPVQETLHGCQPVTQEDGKESRISVQERQ (SEQ ID NO: 72).

[00303] 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.

[00307] In some embodiments, the PSP subunit comprises a transmembrane (TM) domain of a phagocytic receptor.

[00308] In some embodiments, the PSP subunit comprises an ICD domain of a phagocytic receptor.

[00309] In some embodiments, the ICD encoded by the recombinant nucleic acid comprises a domain selected from the group consisting of lectin, dectin 1, mannose receptor (CD206), scavenger receptor A1 (SRA1), MARCO, CD36, CD163, MSR1, SCARA3, COLEC12, SCARA5, SCARB1, SCARB2, CD68, OLR1, SCARF1, SCARF2, CXCL16, STAB1, STAB2, SRCRB4D, SSC5D, CD205, CD207, CD209, RAGE, CD14, CD64, F4/80, CCR2, CX3CR1, CSF1R, Tie2, HuCRIg(L), and CD169 receptor.

[00310] In some embodiments, the ICD comprises the signaling domain derived from any one or more of: lectin, dectin 1, mannose receptor (CD206), scavenger receptor A1 (SRA1), MARCO (Macrophage Receptor with Collagenous Structure, aliases: SRA6, SCARA2), CD36 (Thrombospondin receptor, aliases: Scavenger Receptor class B, member 3), CD163 (Scavenger receptor, cysteine rich- type 1), MSR1, SCARA3, COLEC12 (aliases: Scavenger Receptor With C-Type Lectin, SCARA4, or Collectin 12), SCARA5, SCARB1, SCARB2, CD68 (SCARD, microsialin), OLR1 (Oxidized Low Density Lipoprotein Receptor 1, LOX1, or C-Type Lectin Domain Family 8 Member A), SCARF1, SCARF2, SRCRB4D, SSC5D, and CD169 (aliases, Sialoadhesin receptor, SIGLEC1).

[00311] In some embodiments, the recombinant nucleic acid encodes, for example, an intracellular domain of human MARCO. The PSR subunit comprises an intracellular domain having a 44 amino acid ICD of human MARCO having an amino acid sequence:

MRNKKILKEDELLSETQQAAFHQIAMEPFENVKPKRRNGVNF (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 MARCO.

[00312] In some embodiments, for example, the PSR (phagocytic scavenger receptor) comprises a transmembrane region of human MARCO.

[00313] 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: MEQWDHFHNQQEDTDSCSESVKFDARSMTA

LLPPNPKNSPSLQEKLSFK (SEQ ID NO: 74). 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.

[00314] In some embodiments, the PSR comprises a transmembrane region of human SRA1.

[00315] 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

(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 (SEQ ID NO: 76) (where x is multiple copies of the moiety, designated as 1, 2, 3, 4, and so on) 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.

[00387] 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 acids. In some embodiments, the linker comprises 6 amino acids. In some embodiments, the linker comprises 7 amino acids. In some embodiments, the linker comprises 8 amino acids. In some embodiments, the linker comprises 9 amino acids. In some embodiments, the linker comprises 8 amino acids. In some embodiments, the linker comprises 10 amino acids. In some embodiments, the linker comprises 11 amino acids. In some embodiments, the linker comprises 12 amino acids. In some embodiments, the linker comprises 13 amino acids. In some embodiments, the linker comprises 14 amino acids. In some embodiments, the linker comprises 15 amino acids. In some embodiments, the linker comprises 16 amino acids. In some embodiments, the linker comprises 17 amino acids. In some embodiments, the linker comprises 18 amino acids. In some embodiments, the linker comprises 19 amino acids. In some embodiments, the linker comprises 20 amino acids.

[00388] As contemplated herein, any suitable ECD, TM or ICD domain can be cloned interchangeably in the suitable portion of any one of the CARP receptors described in the disclosure to obtain a protein with enhanced phagocytosis compared to an endogenous receptor.

Characteristics of the Fusion Proteins:

[00389] The CFP can structurally incorporate into the cell membrane of the cell in which it is expressed. Specific leader sequences in the nucleic acid construct, such as the signal peptide can be used to direct plasma membrane expression of the encoded protein. The transmembrane domain encoded by the construct can incorporate the expressed protein in the plasma membrane of the cell.

[00390] In some embodiments, the transmembrane domain comprises a TM domain of an FcRalpha receptor, which dimerizes with endogenous FcR-gamma receptors in the macrophages, ensuring macrophage specific expression.

[00391] The CFP can render the cell that expresses it as potently phagocytic. When the recombinant nucleic acid encoding the CFP is expressed in a cell, the cell can exhibit an increased phagocytosis of a

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

[00482] *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: 62)	Intron encoded homing proteins found in various genera including bacteria
GIY-YIG	This domain is found in the amino terminal region of excinuclease 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.

R2 retroelement DBD
Sso7d
Protein A (ssDNA)
OB-fold (ssDNA)

[00484] VII. Replacing human LINE-1 with LINE-1 from other organisms. In this example, the sequence encoding human LINE-1 is replaced by a LINE-1 from a different organism. In one example, the human LINE-1 construct is compared with a construct where the human LINE-1 is replaced by a minke whale LINE-1 sequence (**FIG. 27**). Using the same experimental framework, a number of ORFs are tested. An exemplary non-exhaustive list is provided in **Table 6** below. A further comprehensive list is available in Ivancevic A. et al., Genome Biol Evol 8(11):3301-3322.

Table 6 - Exemplary LINE-1 elements from organism for use in replacement of the human LINE-1

Species Name	No of total LINE-1/No active/percent active
Balaenoptera acutorostrata scammoni	8,012/5,006/62.4%
Rhinopithecus roxellana	11,115/2,954/26.5%
Mus musculus	18,280/4,143/22.66%
Aedes aegypti	519/184/35.4%
Zea mays	744/165/22.17%
Brassica napus	1,929/565/29.2%
Brassica rapa	543/228/41.9%
Danio rerio	590/268/45.4%

[00485] In another set, human LINE-1 is retained as in the GFP plasmid, but an inhibitor of human LINE-1 silencer is utilized to prevent recognition by endogenous proteins like HUSH complex TASOR protein. In this case, the TASOR inhibitor is an inhibitory RNA, such as a miRNA.

[00486] VIII. LINE-1 fusion proteins for target specificity. In this example, the LINE-1 plasmid GFP ORF2 is fused with a domain of a MegaTAL nuclease, a CRISPR-CAS nuclease, a TALEN, R2 retroelement binding zinc finger binding domain, or a DNA binding domain that can bind to repetitive elements such as Rep78 AAV. **FIG. 28** exemplifies the designs. **Table 7** provides a list of the different elements that can be fused to increase sequence specific retrotransposition.

Table 7 - Exemplary proteins with DNA binding domains to be fused to ORF2 for increasing retrotransposition specificity,

Elements
Transcription Factors

Claims

What is claimed is:

1. A composition comprising a human cell comprising modified genomic DNA, wherein the modified genomic DNA comprises a human insert sequence, wherein the human insert sequence comprises:
 - (a) a DNA sequence encoding an exogenous human therapeutic polypeptide, and
 - (b) a DNA sequence comprising a reverse complement of a human mobile genetic element sequence, wherein the human mobile genetic element sequence is an RNA sequence of the RNA molecule comprising a sequence encoding a polypeptide with target-primed reverse transcription (TPRT) activity, andwherein the human cell expresses the exogenous human therapeutic polypeptide from the human insert sequence.
2. A method of expressing an exogenous therapeutic polypeptide from a genomically integrated DNA sequence of a target human cell population, the method comprising:
 - (a) contacting an RNA molecule to the target human cell population, wherein the RNA molecule comprises:
 - (i) an RNA sequence that is a reverse complement of the DNA sequence encoding the exogenous human therapeutic polypeptide, and
 - (ii) a human mobile genetic element comprising an RNA sequence encoding a polypeptide with target-primed reverse transcription (TPRT) activity; and
 - (b) translating the human mobile genetic element to produce the polypeptide with TPRT activity;
 - (c) reverse transcribing the RNA sequence that is the reverse complement of the DNA sequence encoding the exogenous human therapeutic polypeptide via the TPRT activity of the polypeptide translated in step (b), thereby producing the DNA sequence encoding the exogenous human therapeutic polypeptide;
 - (d) integrating the DNA sequence encoding the exogenous human therapeutic polypeptide into the genomic DNA of a target human cell of the target human cell population at a target site; and
 - (e) expressing the exogenous human therapeutic polypeptide from the genomically integrated DNA sequence encoding the exogenous human therapeutic polypeptide in the target human cell.
3. The composition of claim 1, wherein the human cell is a population of human cells.
4. The method of claim 2, wherein (e) comprises expressing only the exogenous human therapeutic polypeptide from the sequence integrated into the genomic DNA in the target human cell.
5. The composition of claim 3 or the method of any one of claims 2-4, wherein the exogenous human therapeutic polypeptide is expressed in at least 2% of the cells in the human cell population.

6. The composition of claim 3 or the method of any one of claims 2-4, wherein the exogenous human therapeutic polypeptide is expressed in at least 5% of the cells in the human cell population.
7. The composition of claim 3 or the method of any one of claims 2-4, wherein the exogenous human therapeutic polypeptide is expressed in at least 10% of the cells in the human cell population.
8. The composition of any one of claims 1, 3 and 5-7 or the method of any one of claims 2 and 4-7, wherein the RNA molecule is an in vitro transcribed RNA molecule, a synthetic RNA molecule, a purified RNA molecule, or an isolated RNA molecule.
9. The composition of any one of claims 1, 3 and 5-8 or the method of any one of claims 2 and 4-8, wherein the RNA molecule is an mRNA molecule.
10. The composition of any one of claims 1, 3 and 5-9 or the method of any one of claims 2 and 4-9, wherein the polypeptide encoded by the sequence of the human mobile genetic element comprises a human ORF2p or a functional fragment thereof.
11. The composition of claim 10 or the method of claim 10, wherein the RNA molecule comprises a first RNA molecule and a second RNA molecule, wherein the first RNA molecule comprises a sequence encoding a human ORF1p or a functional fragment thereof, and the second RNA molecule encodes the human ORF2p or functional fragment thereof.
12. The method of claim 11, wherein the ratio of the first RNA molecule to the second RNA molecule in the composition is (a) at least 2:1, (b) at most 5:1, or (c) about 3:1.
13. The method of claim 11, wherein the first RNA molecule comprises a first number of copies of a sequence encoding the human ORF1p or functional fragment thereof and the second RNA molecule comprises a second number of copies of a sequence encoding the human ORF2p or functional fragment thereof, wherein the ratio of the first number of copies to the second number of copies is at least 2:1.
14. The composition of any one of claims 10 or 11 or the method of any one of claims 10-13, wherein the polypeptide encoded by the sequence of the human mobile genetic element comprises a sequence with at least 90% sequence identity to SEQ ID NO: 55.
15. The composition of any one of claims 1, 3, 5-11 and 14 or the method of any one of claims 2 and 4-14, wherein the polypeptide encoded by the sequence of the human mobile genetic element comprises a nuclear localization signal (NLS).
16. The composition of any one of claims 10, 11, 14 and 15 or the method of any one of claims 10-15, wherein the composition further comprises a sequence encoding a human ORF1p or a functional fragment thereof, wherein the human ORF1p or functional fragment thereof and the human ORF2p or functional fragment thereof are translated from different open reading frames (ORFs) of the same RNA molecule.
17. The composition of any one of claims 1, 3, 5-11 and 14-16 or the method of any one of claims 2 and 4-16, wherein the human cell is an immune cell selected from the group consisting of a T cell, a B cell, a myeloid cell, a monocyte, a macrophage and a dendritic cell.

18. The composition of any one of claims 1, 3, 5-11 and 14-17 or the method of any one of claims 2 and 4-17, wherein the RNA molecule comprises homology arms complementary to the target site in the genomic DNA.
19. The method of any one of claims 2 and 4-18, wherein integrating comprises integrating the sequence encoding the exogenous therapeutic polypeptide into non-ribosomal genomic DNA of the target human cell or integrating the sequence encoding the exogenous therapeutic polypeptide into the genomic DNA at a locus that is not a ribosomal DNA (rDNA) locus.
20. The composition of any one of claims 1, 3, 5-11, 14-17 and 19 or the method of any one of claims 2 and 4-19, wherein the RNA molecule has a total length of from 3 kb to 20 kb.
21. The composition of any one of claims 1, 3, 5-11, 14-17, 19 and 20 or the method of any one of claims 2 and 4-20, wherein the exogenous human therapeutic polypeptide is a ligand, an antibody, a receptor, an enzyme, a transport protein, a structural protein, a hormone, a contractile protein, a storage protein, a transcription factor, a chimeric antigen receptor (CAR) or a T cell receptor (TCR).
22. The composition of any one of claims 1, 3, 5-11, 14-17, and 19-21 or the method of any one of claims 2 and 4-21, wherein the sequence encoding the exogenous human therapeutic polypeptide does not comprise introns.
23. The composition of any one of claims 1, 3, 5-11, 14-17, and 19-22 or the method of any one of claims 2 and 4-22, wherein the RNA molecule comprises a 5' UTR sequence, a 3' UTR sequence and a poly A sequence; wherein:
 - (i) the 5' UTR sequence is upstream of the sequence of the human mobile genetic element encoding a polypeptide,
 - (ii) the 3' UTR sequence is downstream of the sequence encoding the exogenous therapeutic polypeptide; and
 - (iii) the 3' UTR is upstream of the poly A sequence; andwherein the 5' UTR sequence, the 3' UTR sequence or the poly A sequence comprises a binding site for the human ORF2p or a functional fragment thereof.
24. The composition of any one of claims 1, 3, 5-11, 14-17, and 19-23 or the method of any one of claims 2 and 4-23, wherein the RNA molecule comprises a 5' UTR sequence and a 3' UTR sequence, wherein
 - (a) the 5' UTR comprises a 5' UTR from LINE-1; and/or
 - (b) the 3' UTR comprises a 3' UTR from LINE-1.
25. The composition of any one of claims 1, 3, 5-11, 14-17, and 19-24 or the method of any one of claims 2 and 4-24, wherein the RNA molecule comprises a sequence encoding a nuclease domain, wherein the nuclease domain is a nuclease domain from megaTAL, TALEN, Cas9, Cas6, Cas7, or Cas8; wherein the nuclease domain is not from ORF2p.
26. The composition of any one of claims 10, 11, 14-17, and 19-25 or the method of any one of claims 10-25, wherein the human ORF2p or functional fragment thereof comprises a mutation that reduces

endonuclease activity of the ORF2p or functional fragment thereof compared to the endonuclease activity of an ORF2p or functional fragment thereof without the mutation, wherein the mutation is S228P.

27. A composition comprising an isolated or purified RNA molecule for use as a medicament, wherein the RNA molecule comprises
 - (i) an RNA sequence that is a reverse complement of the DNA sequence encoding the exogenous human therapeutic polypeptide, and
 - (ii) a human mobile genetic element comprising an RNA sequence encoding a polypeptide with target-primed reverse transcription (TPRT) activity.
28. Use of a composition comprising an isolated or purified RNA molecule in the manufacture of a medicament for treating a disease or condition in a human subject, wherein the RNA molecule comprises
 - (i) an RNA sequence that is a reverse complement of a DNA sequence encoding an exogenous human therapeutic polypeptide, and
 - (ii) a human mobile genetic element comprising an RNA sequence encoding a polypeptide with target-primed reverse transcription (TPRT) activity.