SYSTEMS, METHODS, AND COMPOSITIONS FOR SITE-SPECIFIC GENETIC ENGINEERING USING PROGRAMMABLE ADDITION VIA SITE-SPECIFIC TARGETING ELEMENTS (PASTE)

CROSS REFERENCE TO RELATED APPLICATIONS

[1] This application claims the benefit of U.S. Provisional Patent Application Serial No. 63/222,550, filed July 16, 2021 and U.S. Provisional Patent Application Serial No. 63/094,803, filed October 21, 2020. The entire contents of the above-referenced patent applications are incorporated by reference in their entirety herein.

FIELD OF DISCLOSURE

[2] The subject matter disclosed herein is generally directed to systems, methods, and compositions for site-specific genetic engineering using Programmable Addition via Site-Specific Targeting Elements (PASTE) for the treatment of diseases and diagnostics.

BACKGROUND

[3] Editing genomes using the RNA-guided DNA targeting principle of CRISPR-Cas (Clustered Regularly Interspaced Short Palindromic Repeats-CRISPR associated proteins) immunity has been widely exploited and has become a powerful genome editing means for a wide variety of applications. The main advantage of CRISPR-Cas system lies in the minimal requirement for programmable DNA interference: an endonuclease, such as a Cas9, Cas12, or any programmable nucleases, guided by a customizable dual-RNA structure. Cas9 is a multi-domain enzyme that uses an HNH nuclease domain to cleave the target strand. The CRISPR/Cas9 protein-RNA complex is localized on the target by a guide RNA (guide RNA), then cleaved to generate a DNA double strand break (dsDNA break, DSB). After cleavage, DNA repair mechanisms are activated to repair the cleaved strand. Repair mechanisms are generally from one of two types: non-homologous end joining (NHEJ) or homologous recombination (HR). In general, NHEJ dominates the repair, and, being error prone, generates random indels (insertions or deletions) causing frame shift mutations, among others. In contrast, HR has a more precise repairing capability and is potentially capable of incorporating the exact substitution or insertion. To enhance HR, several techniques have been tried, for example: combination of fusion proteins of Cas9 nuclease with homology-directed repair (HDR) effectors to enforce their localization at DSBs, introducing an

overlapping homology arm, or suppression of NHEJ. Most of these techniques rely on the host DNA repair systems.

[4] Recently, new guided editors have been developed, such as guided prime editors (PE) PE1, PE2, and PE3, e.g., Liu, D. *et al., Nature* 2019, 576, 149-157. These PEs are reverse transcriptase (RT) fused with Cas 9 H 840A nickase (Cas9n (H840A)), and the genome editing is achieved using a prime-editing guide RNA (pegRNA). Despite these developments, programmable gene integration is still generally dependent on cellular pathways or repair processes.

[5] Therefore, there is a need for more effective tools for gene editing and delivery.

SUMMARY

[6] The present disclosure provides a method of site-specific integration of a nucleic acid into a cell genome. The method comprises incorporating an integration site at a desired location in the cell genome by introducing into the cell a DNA binding nuclease linked to a reverse transcriptase domain, wherein the DNA binding nuclease comprises a nickase activity; and a guide RNA (gRNA) comprising a primer binding sequence linked to an integration sequence, wherein the gRNA interacts with the DNA binding nuclease and targets the desired location in the cell genome, wherein the DNA binding nuclease nicks a strand of the cell genome and the reverse transcriptase domain incorporates the integration sequence of the gRNA into the nicked site, thereby providing the integration site at the desired location of the cell genome. The method further comprises integrating the nucleic acid into the cell genome by introducing into the cell a DNA or RNA strand comprising the nucleic acid linked to a sequence that is complementary or associated to the integration site, and an integration enzyme, wherein the integration enzyme incorporates the nucleic acid into the cell genome at the integration site by integration, recombination, or reverse transcription of the sequence that is complementary or associated to the integration site, thereby introducing the nucleic acid into the desired location of the cell genome of the cell.

[7] In some embodiments, the gRNA can be hybridized to a complementary strand of the cell genome to the genomic strand that is nicked by the DNA binding nuclease.

[8] In some embodiments, the integration enzyme can be introduced as a peptide or a nucleic acid encoding the same.

[9] In some embodiments, the DNA binding nuclease can be introduced as a peptide or a nucleic acid encoding the same.

[10] In some embodiments, the DNA or RNA strand comprising the nucleic acid can be introduced into the cell as a minicircle, a plasmid, mRNA or a linear DNA.

[11] In some embodiments, the DNA or RNA strand comprising the nucleic acid can be between 1000 bp and 10,000 bp.

[12] In some embodiments, the DNA or RNA strand comprising the nucleic acid can be more than 10,000 bp.

[13] In some embodiments, the DNA or RNA strand comprising the nucleic acid can be less than 1000 bp.

[14] In some embodiments, the DNA comprising the nucleic acid can be introduced into the cell as a minicircle.

[15] In some embodiment, the minicircle cannot comprise sequences of a bacterial origin.

[16] In some embodiments, the DNA binding nuclease can be linked to a reverse transcriptase domain and the integration enzyme can be linked via a linker. The linker can be cleavable. The linker can be non-cleavable. The linker can be replaced by two associating binding domains of the DNA binding nuclease linked to a reverse transcriptase.

[17] In some embodiments, the integration enzyme can be selected from the group consisting of Cre, Dre, Vika, Bxb1, φC31, RDF, FLP, φBT1, R1, R2, R3, R4, R5, TP901-1, A118, φFC1, φC1, MR11, TG1, φ370.1, Wβ, BL3, SPBc, K38, Peaches, Veracruz, Rebeuca, Theia, Benedict, KSSJEB, PattyP, Doom, Scowl, Lockley, Switzer, Bob3, Troube, Abrogate, Anglerfish, Sarfire, SkiPole, ConceptII, Museum, Severus, Airmid, Benedict, Hinder, ICleared, Sheen, Mundrea, BxZ2, φRV, retrotransposases encoded by R2, L1, Tol2 Tc1, Tc3, Mariner (Himar 1), Mariner (mos 1), and Minos, and any mutants thereof.

[18] In some embodiments, the integration enzyme can be Bxb1 or a mutant thereof.

[19] In some embodiments, the integration site can be selected from an attB site, an attP site, an attL site, an attR site, a lox71 site a Vox site, or a FRT site.

[20] In some embodiments, the DNA binding nuclease comprising a nickase activity can be selected from Cas9-D10A, Cas9-H840A, and Cas12a/b nickase.

[21] In some embodiments, the reverse transcriptase domain can be selected from the group consisting of Moloney Murine Leukemia Virus (M-MLV) reverse transcriptase domain, transcription xenopolymerase (RTX), avian myeloblastosis virus reverse transcriptase (AMV-RT), and Eubacterium rectale maturase RT (MarathonRT).

[22] In some embodiments, the reverse transcriptase domain can comprise a mutation relative to the wild-type sequence.

[23] In some embodiments, the M-MLV reverse transcriptase domain can comprise one or more mutations selected from the group consisting of D200N, T306K, W313F, T330P and L603W.

[24] In some embodiments, the method can further comprise introducing a second nicking guide RNA (ngRNA). The ngRNA can direct nicking at 90 bases downstream of the gRNA nick on a complementary strand.

[25] In some embodiments, the gRNA, the nucleic acid encoding the DNA binding nuclease, the reverse transcriptase, the DNA comprising nucleic acid linked to a complementary integration site, the integration enzyme, and optionally the ngRNA can be introduced into a cell in a single reaction.

[26] In some embodiments, the gRNA, the nucleic acid encoding the DNA binding nuclease, the reverse transcriptase, the DNA comprising nucleic acid linked to a complementary integration site, the integration enzyme, and optionally the ngRNA can be introduced using a virus, a RNP, an mRNA, a lipid, or a polymeric nanoparticle.

[27] In some embodiments, the nucleic acid can be a reporter gene. The reporter gene can be a fluorescent protein.

[28] In some embodiments, the cell can be a dividing cell.

[29] In some embodiments, the cell can be a non-dividing cell.

[30] In some embodiments, the desired location in the cell genome can be the locus of a mutated gene.

[31] In some embodiments, the nucleic acid can be a degradation tag for programmable knockdown of proteins in the presence of small molecules.

[32] In some embodiments, the cell can be a mammalian cell, a bacterial cell or a plant cell.

[33] In some embodiments, nucleic acid can be a T-cell receptor (TCR), a chimeric antigen receptor (CAR), an interleukin, a cytokine, or an immune checkpoint gene for integration into a T-cell or natural killer (NK) cell. The TCR, the CAR, the interleukin, the cytokine, or the immune checkpoint gene can be incorporated into the target site of the T-cell or NK cell genome using a minicircle DNA.

[34] In some embodiments, the nucleic acid can be a beta hemoglobin (HBB) gene and the cell can be a hematopoietic stem cell (HSC). The HBB gene can be incorporated into the target site in the HSC genome using a minicircle DNA. The nucleic acid can be a gene responsible for beta thalassemia or sickle cell anemia.

[35] In some embodiments, the nucleic acid can be a metabolic gene. The metabolic gene can be involved in alpha-1 antitrypsin deficiency or ornithine transcarbamylase (OTC) deficiency. The metabolic gene can be a gene involved in inherited diseases.

[36] In some embodiments, the nucleic acid can be a gene involved in an inherited disease or an inherited syndrome. The inherited disease can be cystic fibrosis, familial hypercholesterolemia, adenosine deaminase (ADA) deficiency, X-linked SCID (X-SCID), Wiskott-Aldrich syndrome (WAS), hemochromatosis, Tay-Sachs, fragile X syndrome, Huntington's disease, Marfan syndrome, phenylketonuria, or muscular dystrophy.

[37] The present disclosure provides a vector comprising a nucleic acid encoding a DNA binding nuclease comprising a nickase activity C-terminally linked to a reverse transcriptase linked to an integration enzyme via a linker.

[38] In some embodiments, the linker can be cleavable.

[39] In some embodiments, the linker can be non-cleavable.

[40] In some embodiments, the linker can comprise two associating binding domains of the DNA binding nuclease linked to a reverse transcriptase.

[41] In some embodiments, the integration enzyme can comprise a conditional activation domain or conditional expression domain.

[42] In some embodiments, the integration enzyme can be fused to an estrogen receptor.

[43] In some embodiments, the DNA binding nuclease comprising a nickase activity can be selected from the group consisting of Cas9-D10A, Cas9-H840A, and Cas12a/b.

[44] In some embodiments, the reverse transcriptase can be a M-MLV reverse transcriptase, a AMV-RT, MarathonRT, or a RTX. The reverse transcriptase can be a modified M-MLV reverse transcriptase relative to the wildtype M-MLV reverse transcriptase. The M-MLV reverse transcriptase domain can comprise one or more of the mutations selected from the group consisting of D200N, T306K, W313F, T330P and L603W.

[45] In some embodiments, the integration enzyme can be selected from the group consisting of Cre, Dre, Vika, Bxb1, φC31, RDF, FLP, φBT1, R1, R2, R3, R4, R5, TP901-1, A118, φFC1, φC1, MR11, TG1, φ370.1, Wβ, BL3, SPBc, K38, Peaches, Veracruz, Rebeuca, Theia, KSSJEB, PattyP, Doom, Scowl, Lockley, Switzer, Bob3, Troube, Abrogate, Anglerfish, Sarfire, SkiPole, ConceptII, Museum, Severus, Airmid, Benedict, Hinder, ICleared, Sheen, Mundrea, BxZ2, φRV, retrotransposases encoded by R2, L1, Tol2 Tc1, Tc3, Mariner (Himar 1), Mariner (mos 1), and Minos, and any mutants thereof.

[46] In some embodiments, the recombinase or integrase can be Bxb1 or a mutant thereof.

[47] The present disclosure provides a cell comprising a vector comprising a nucleic acid encoding a DNA binding nuclease comprising a nickase activity C-terminally linked to a reverse transcriptase linked to an integration enzyme via a linker. The cell further comprises a gRNA comprising a primer binding sequence, an integration sequence, and a guide sequence, wherein the gRNA can interact with the encoded nuclease comprising a nickase activity. The cell further comprising a DNA minicircle comprising a nucleic acid and a sequence recognized by the encoded integrase, recombinase, or reverse transcriptase. The cell further comprising a nicking guide RNA (ngRNA) capable of binding the encoded nuclease comprising a nickase activity, and wherein the ngRNA targets a sequence away from the gRNA.

[48] In some embodiments, the minicircle cannot comprise a sequence of bacterial origin.

[49] In some embodiments, the integration enzyme can be selected from the group consisting of Cre, Dre, Vika, Bxb1, φC31, RDF, FLP, φBT1, R1, R2, R3, R4, R5, TP901-1, A118, φFC1, φC1, MR11, TG1, φ370.1, Wβ, BL3, SPBc, K38, Peaches, Veracruz, Rebeuca, Theia, KSSJEB, PattyP, Doom, Scowl, Lockley, Switzer, Bob3, Troube, Abrogate, Anglerfish, Sarfire, SkiPole, ConceptII, Museum, Severus, Airmid, Benedict, Hinder, ICleared, Sheen, Mundrea, BxZ2, φRV, retrotransposases encoded by R2, L1, Tol2 Tc1, Tc3, Mariner (Himar 1), Mariner (mos 1), and Minos, and any mutants thereof.

[50] In some embodiments, the integration enzyme can be Bxb1 or a mutant thereof.

[51] In some embodiments, the DNA binding nuclease comprising a nickase activity can be selected from the group consisting of Cas9-D10A, Cas9-H840A and Cas12a.

[52] In some embodiments, the reverse transcriptase can be a M-MLV reverse transcriptase. The reverse transcriptase can be a modified M-MLV reverse transcriptase. The amino acid sequence of the M-MLV reverse transcriptase can comprise one or more mutations selected from the group consisting of D200N, T306K, W313F, T330P, and L603W.

[53] In some embodiments, the cell can further comprise introducing ngRNA to the cell. The ngRNA can be a +90 ngRNA. The +90 ngRNA can direct nicking at 90 bases downstream of the gRNA nick on a complementary strand.

[54] The present disclosure provides a polypeptide comprising a DNA binding nuclease comprising a nickase activity C-terminally linked to a reverse transcriptase linked to an integration enzyme via a linker.

[55] In some embodiments, the linker can be cleavable.

[56] In some embodiments, the linker can be non-cleavable.

[57] In some embodiments, the integration enzyme can be fused to an estrogen receptor.

[58] In some embodiments, the DNA binding nuclease comprising a nickase activity can be selected from the group consisting of Cas9-D10A, Cas9-H840A, and Cas12a/b.

[59] In some embodiments, the reverse transcriptase can be a M-MLV reverse transcriptase, a AMV-RT, a MarathonRT, or a XRT. The reverse transcriptase can be a modified M-MLV relative to a wild-type M-MLV reverse transcriptase. The M-MLV reverse transcriptase domain can comprise one or more of mutations selected from the group consisting of D200N, T306K, W313F, T330P, and L603W.

[60] In some embodiments, the integration enzyme can be selected from group consisting of Cre, Dre, Vika, Bxb1, φC31, RDF, FLP, φBT1, R1, R2, R3, R4, R5, TP901-1, A118, φFC1, φC1, MR11, TG1, φ370.1, Wβ, BL3, SPBc, K38, Peaches, Veracruz, Rebeuca, Theia, KSSJEB, PattyP, Doom, Scowl, Lockley, Switzer, Bob3, Troube, Abrogate, Anglerfish, Sarfire, SkiPole, ConceptII, Museum, Severus, Airmid, Benedict, Hinder, ICleared, Sheen, Mundrea, BxZ2, φRV, retrotransposases encoded by R2, L1, Tol2 Tc1, Tc3, Mariner (Himar 1), Mariner (mos 1), and Minos, and any mutants thereof.

[61] In some embodiments, the integration enzyme can be Bxb1 or a mutant thereof.

[62] The present disclosure provides a gRNA that specifically binds to a DNA binding nuclease comprising nickase activity, the gRNA comprising a primer binding site, which hybridizes to a nicked DNA strand, a recognition site for an integration enzyme, and a target recognition sequence recognizing a target site in a cell genome and hybridizing to a genomic strand complementary to the strand that is nicked by the DNA binding nuclease.

[63] In some embodiments, the DNA binding nuclease comprising a nickase activity can be selected from the group consisting of Cas9-D10A, Cas9-H840A, and Cas12a/b.

[64] In some embodiments, the primer binding site can hybridize to the 3′ end of the nicked DNA strand.

[65] In some embodiments, the recognition site for the integration enzyme can be selected from an attB site, an attP site, an attL site, an attR site, a lox71 site, and a FRT site.

[66] In some embodiments, the recognition site for the integration enzyme can be a Bxb1 site.

[67] The present disclosure provides a method of site-specific integration of two or more nucleic acids into a cell genome. The method comprises incorporating two integration sites at desired locations in the cell genome by introducing into the cell a DNA binding nuclease

linked to a reverse transcriptase domain, wherein the DNA binding nuclease comprises a nickase activity, and two guide RNAs (gRNAs), each comprising, a primer binding sequence, linked to a unique integration sequence, wherein the gRNA interacts with the DNA binding nuclease and targets the desired locations in the cell genome, wherein the DNA binding nuclease nicks a strand of the cell genome and the reverse transcriptase domain incorporates each of the integration sequence of the gRNA into the nicked site, thereby providing the integration site at the desired locations of the cell genome. The method further comprises integrating the nucleic acid by introducing into the cell two or more DNA or RNA comprising the nucleic acids, wherein each DNA is flanked by orthogonal integration sites, and an integration enzyme, wherein the integration enzyme incorporates the nucleic acids into the cell genome at the integration sites by integrase, recombinase, or reverse transcriptase of the sequence that is complementary or associated to the integration site, thereby introducing the nucleic acids into the desired locations of the cell genome of the cell.

[68] In some embodiments, each of the two different integration sites inserted into the cell genome can be attB sequences comprising different palindromic or non-palindromic central dinucleotide.

[69] In some embodiments, each of the two different integration sites inserted into the cell genome can be attP sequences comprising different palindromic or non-palindromic central dinucleotide.

[70] In some embodiments, the integration enzyme can enable each of the two or more DNA or RNA comprising the nucleic acids to directionally enable integration of the nucleic acids into a genome via recombination of a pair of orthogonal attB site sequence and an attP site sequence.

[71] In some embodiments, the integration enzyme can be selected from the group consisting of Cre, Dre, Vika, Bxb1, φC31, RDF, FLP, φBT1, TP901-1, A118, φFC1, φC1, MR11, TG1, φ370.1, Wβ, BL3, SPBc, K38, Peaches, Veracruz, Rebeuca, Theia, KSSJEB, PattyP, Doom, Scowl, Lockley, Switzer, Bob3, Troube, Abrogate, Anglerfish, Sarfire, SkiPole, ConceptII, Museum, Severus, Airmid, Benedict, Hinder, ICleared, Sheen, Mundrea, BxZ2, φRV, retrotransposases encoded by R1, R2, R3, R4, R5, L1, Tol2 Tc1, Tc3, Mariner (Himar 1), Mariner (mos 1), and Minos, and any mutants thereof.

[72] In some embodiments, the integration enzyme can be Bxb1 or a mutant thereof.

[73] In some embodiments, the DNA comprising genes can be genes involved in a cell maintenance pathway, cell-division, or a signal transduction pathway.

[74] In some embodiments, the reverse transcriptase domain can comprise Moloney Murine Leukemia Virus (M-MLV) reverse transcriptase domain, transcription xenopolymerase (RTX), avian myeloblastosis virus reverse transcriptase (AMV-RT), or Eubacterium rectale maturase RT (MarathonRT).

[75] In some embodiments, the DNA binding nuclease comprising a nickase activity can be selected from the group consisting of Cas9-D10A, Cas9-H840A, and Cas12a/b.

[76] In some embodiments, the pair of an attB site sequence and an attP site sequence can be selected from the group consisting of SEQ ID NO: 5 and SEQ ID NO: 6, SEQ ID NO: 7 and SEQ ID NO: 8, SEQ ID NO: 9 and SEQ ID NO: 10, SEQ ID NO: 11 and SEQ ID NO: 12, SEQ ID NO: 13 and SEQ ID NO: 14, SEQ ID NO: 15 and SEQ ID NO: 16, SEQ ID NO: 17 and SEQ ID NO: 18, SEQ ID NO: 19 and SEQ ID NO: 20, SEQ ID NO: 21 and SEQ ID NO: 22, SEQ ID NO: 23 and SEQ ID NO: 24, SEQ ID NO: 25 and SEQ ID NO: 26, SEQ ID NO: 27 and SEQ ID NO: 28, SEQ ID NO: 29 and SEQ ID NO: 30, SEQ ID NO: 31 and SEQ ID NO: 32, SEQ ID NO: 33 and SEQ ID NO: 34 and SEQ ID NO: 35 and SEQ ID NO: 36.

[77] The present disclosure provides a cell comprising a vector comprising a nucleic acid encoding a DNA binding nuclease comprising a nickase activity, wherein the DNA binding nuclease is C-terminally linked to a reverse transcriptase, wherein the reverse transcriptase is linked to a recombinase or integrase via a linker. The cell further comprises two guide RNAs (gRNAs) comprising a primer binding sequence, an integration sequence and a guide sequence, wherein the gRNA can interact with the encoded DNA binding nuclease comprising a nickase activity. The cell further comprises two or more DNA or RNA strands comprising a nucleic acid and a pair of flanking attB site sequence and an attP site sequence recognized by the encoded integrase or recombinase. The cell optionally further comprises a nicking guide RNA (ngRNA) capable of binding the encoded nuclease comprising a nickase activity, and wherein the ngRNA targets a sequence away from the gRNA.

[78] The present disclosure provides a cell comprising a modified genome, wherein the modification comprises incorporation of two orthogonal integration sites within the cell genome by introducing into the cell a: vector comprising a nucleic acid encoding a DNA binding nuclease comprising a nickase activity, wherein the DNA binding nuclease is C-terminally linked to a reverse transcriptase; two guide RNAs (gRNAs), each comprising a primer binding sequence, a genomic integration sequence, and a guide sequence, wherein the gRNA can interact with the encoded nuclease comprising a nickase activity; and optionally a

nicking guide RNA (ngRNA) capable of binding the encoded nuclease comprising a nickase activity, and wherein the ngRNA targets a sequence away from the gRNA.

[79] The present disclosure provides a method of integrating two or more nucleic acids into the cell genome of cell of claim 90, the method comprising introducing into the cell: two or more DNA, each comprising a nucleic acid and a pair of flanking orthogonal integration site sequences; an integration enzyme that can recognize the integration site sequence enabling directional linking of the two or more DNA comprising nucleic acid; and enabling incorporation of the nucleic acids into the cell genome by integrating the 5' orthogonal integration sequence of the first DNA with the first genomic integration sequence and 3' orthogonal integration sequence of the last DNA with the last genomic integration sequence, thereby incorporating the two or more nucleic acids into the cell genome.

[80] The present disclosure provides a cell comprising a modified genome, wherein the modification comprises incorporation of two orthogonal integration sites within the cell genome by introducing into the cell: a vector comprising a nucleic acid encoding a DNA binding nuclease comprising a nickase activity, wherein the DNA binding nuclease is C-terminally linked to a reverse transcriptase; two guide RNAs (gRNAs), each comprising a primer binding sequence, a genomic integration sequence, and a guide sequence, wherein the gRNA can interact with the encoded nuclease comprising a nickase activity; and optionally a nicking guide RNA (ngRNA) capable of binding the encoded nuclease comprising a nickase activity, and wherein the ngRNA targets a sequence away from the gRNA; two or more DNA or RNA comprising the nucleic acids, wherein each DNA is flanked by orthogonal integration sites; and an integration enzyme, wherein the integration enzyme incorporates the nucleic acids into the cell genome at the integration sites.

BRIEF DESCRIPTION OF THE DRAWINGS

[81] Aspects, features, benefits and advantages of the embodiments described herein will be apparent with regard to the following description, appended claims, and accompanying drawings where:

[82] **FIG. 1** shows a schematic diagram of a concept of Programmable Addition via Site-Specific Targeting Elements (PASTE) according to embodiments of the present teachings;

[83] **FIG. 2** shows a schematic diagram of a prime editing process according to embodiments of the present teachings;

[84] **FIG. 3** shows the percent integration of green fluorescent protein (GFP) in the lentiviral integrated lox71 site in HEK293FT cell line in the presence of various plasmids according to embodiments of the present teachings;

[85] **FIG. 4** shows the percent editing of the HEK293FT genome for incorporation of various lengths of lox71 or lox66 according to embodiments of the present teachings;

[86] **FIG. 5A** shows the percent editing of lox71 site with different PE/Cre vectors according to embodiments of the present teachings;

[87] **FIG. 5B** shows the percent integration of GFP at the lox71 site in HEK293FT cell genome according to embodiments of the present teachings;

[88] **FIG. 6** shows a schematic representation of using Bxb1 to integrate a nucleic acid into the genome according to embodiments of the present teachings;

[89] **FIG. 7** shows the percent integration of GFP or Gluc into the attB locus using Bxb1 Programmable Addition via Site-Specific Targeting Elements (PASTE) according to embodiments of the present teachings;

[90] **FIG. 8** shows the percent editing of various HEK3 targeting pegRNA Programmable Addition via Site-Specific Targeting Elements (PASTE) according to embodiments of the present teachings;

[91] **FIG. 9A** shows a fluorescent image of cells wherein the SUPT16H marker is tagged with EGFP using PASTE according to embodiments of the present teachings;

[92] **FIG. 9B** shows a fluorescent image of cells wherein the SRRM2 marker is tagged with EGFP using Programmable Addition via Site-Specific Targeting Elements (PASTE) according to embodiments of the present teachings;

[93] **FIG. 9C** shows a fluorescent image of cells wherein the LAMNB1 marker is tagged with EGFP using Programmable Addition via Site-Specific Targeting Elements (PASTE) according to embodiments of the present teachings;

[94] **FIG. 9D** shows a fluorescent image of cells wherein the NOLC1 marker is tagged with EGFP using Programmable Addition via Site-Specific Targeting Elements (PASTE) according to embodiments of the present teachings;

[95] **FIG. 9E** shows a fluorescent image of cells wherein the NOLC1 marker is tagged with EGFP using Programmable Addition via Site-Specific Targeting Elements (PASTE) according to embodiments of the present teachings;

[96] **FIG. 9F** shows a fluorescent image of cells wherein the NOLC1 marker is tagged with EGFP using Programmable Addition via Site-Specific Targeting Elements (PASTE) according to embodiments of the present teachings;

[97] **FIG. 9G** shows a fluorescent image of cells wherein the DEPDC4 marker is tagged with EGFP using Programmable Addition via Site-Specific Targeting Elements (PASTE) according to embodiments of the present teachings;

[98] **FIG. 10A** shows comparisons of lipofectamine aided transfection in blue with electroporation aided transfection in red for the addition of the Bxb1 attB site at the ACTB N-terminal site in the genome using PASTE according to embodiments of the present teachings;

[99] **FIG. 10B** shows comparisons of lipofectamine aided transfection in blue with electroporation aided transfection in red for EGFP integration at the ACTB N-terminal site in the genome using PASTE according to embodiments of the present teachings;

[100] **FIG. 11** shows a diagram of the integration of EGFP and Gluc with various HEK3 targeting pegRNAs according to embodiments of the present teachings;

[101] **FIG. 12** shows a schematic diagram of the using φC31 as the integration enzyme, according to embodiments of the present teachings;

[102] **FIG. 13** shows a schematic diagram of multiplexing involving inserting multiple genes of interest in multiple loci using unique guide RNAs that incorporated exterior flanking attB sites according to embodiments of the present teachings;

[103] **FIG. 14A** shows a diagram of the orthogonal editing with the right GT-EGFP according to embodiments of the present teachings;

[104] **FIG. 14B** shows a diagram of the orthogonal editing with the right GA-mCherry according to embodiments of the present teachings;

[105] **FIG. 15A** shows a fluorescent image of a multiplexing of ACTB-EGFP and NOLC1-mCherry according to embodiments of the present teachings

[106] **FIG. 15B** shows a fluorescent image of a multiplexing of ACTB-EGFP and LAMNB1-mCherry according to embodiments of the present teachings;

[107] **FIG. 16A** shows next generation sequencing results of 9X9 attP and attB central dinucleotide variants and their edit percentage wherein the orthogonality of attB/attP combinations for potential multiplexing applications is shown according to embodiments of the present teachings;

[108] **FIG. 16B** shows an heatmap of 9X9 attP and attB central dinucleotide variants and their edit percentage according to embodiments of the present teachings;

[109] **FIG. 17** shows integration of SERPINA and CPS1 into Albumin loci using Albumin guide-pegRNA in HEK293FT cells according to embodiments of the present teachings;

[110] **FIG. 18** shows schematics for different nucleic acids for engineering T-cells according to embodiments of the present teachings;

[111] **FIG. 19** shows the editing efficiency for EGFP integration at the ACTB locus in primary T-cells according to embodiments of the present teachings;

[112] **FIG. 20** shows editing in TRAC locus in HEK293FT with different pegRNA according to embodiments of the present teachings;

[113] **FIG. 21A** shows the attB integration at the ACTB locus using nicking guides 1 and 2 according to embodiments of the present teachings;

[114] **FIG. 21B** shows the EGFP integration at the ACTB locus using nicking guides 1 and 2 according to embodiments of the present teachings;

[115] **FIG. 21C** shows the EGFP integration at an ACTB site according to embodiments of the present teachings;

[116] **FIG. 22A** shows PASTE editing in liver hepatocellular carcinoma cell line HEPG2 according to embodiments of the present teachings;

[117] **FIG. 22B** shows PASTE editing of chronic myelogenous leukemia cell line K562 according to embodiments of the present teachings;

[118] **FIG. 23A** shows the attB addition with targeting and non-targeting guides according to embodiments of the present teachings;

[119] **FIG. 23B** shows the EGFP integration with targeting and non-targeting guides according to embodiments of the present teachings;

[120] **FIG. 23C** shows the EGFP integration for mutagenized Bxb1 according to embodiments of the present teachings;

[121] **FIG. 24A** shows a schematic of the design parameters for the pegRNA according to embodiments of the present teachings;

[122] **FIG. 24B** shows a schematic of the design parameters for nicking guide RNA according to embodiments of the present teachings;

[123] **FIG. 25A** shows the integration of EGFP at the ACTD locus with different PBS and RT lengths according to embodiments of the present teachings;

[124] **FIG. 25B** shows the integration of EGFP at the LMNB1 loci with different PBS and RT lengths according to embodiments of the present teachings;

[125] **FIG. 25C** shows the integration of EGFP at the NOLC1 loci with different PBS and RT lengths according to embodiments of the present teachings;

[126] **FIG. 25D** shows the integration of EGFP at the GRSF1 locus with different PBS and RT lengths and different nicking guides according to embodiments of the present teachings;

[127] **FIG. 25E** shows EGFP integration with mutant attP sites according to embodiments of the present teachings;

[128] **FIG. 25F** shows the PASTE editing of an expanded panel of genes according to embodiments of the present teachings;

[129] **FIG. 26A** shows the PASTE EGPF editing at the ACTB locus according to embodiments of the present teachings;

[130] **FIG. 26B** shows the HITI EGPF editing at the ACTB locus according to embodiments of the present teachings;

[131] **FIG. 26C** shows the comparison between the PASTE and HITI editing a panel of 14 genes according to embodiments of the present teachings;

[132] **FIG. 26D** shows PASTE Bxb1 off-target integrations according to embodiments of the present teachings;

[133] **FIG. 26E** shows PASTE Cas9 off-target integrations according to embodiments of the present teachings;

[134] **FIG. 26F** shows the EGFP integration for gene inserts of different sizes according to embodiments of the present teachings;

[135] **FIG. 27A** shows the orthogonality between selected sets of attB and attP sites according to embodiments of the present teachings;

[136] **FIG. 27B** shows the orthogonality between selected sets of attB and attP sites according to embodiments of the present teachings;

[137] **FIG. 27C** shows a schematic for the orthogonal PASTE editing using engineered di-nucleotide combinations according to embodiments of the present teachings;

[138] **FIG. 28A** shows fluorescent images of the GFP tagging of ACTB and SUPT16H genes with PASTE according to embodiments of the present teachings;

[139] **FIG. 28B** shows fluorescent images of the GFP tagging of NOLC1 and SRRM2 genes with PASTE according to embodiments of the present teachings;

[140] **FIG. 28C** shows fluorescent images of the GFP tagging of LMNB1 and DEPDC4 genes with PASTE according to embodiments of the present teachings;

[141] **FIG. 28D** shows the orthogonal gene integration at three endogenous sites with PASTE according to embodiments of the present teachings;

[142] **FIG. 28E** shows the multiplexed insertion via one-plex, two-plex, and three-plex gene insertion at three endogenous sites via PASTE according to embodiments of the present teachings;

[143] **FIG. 28F** shows fluorescent images of two single cells with multiplexed gene tagging of ACTB (EGFP) and NOLC1 (mCherry) using PASTE according to embodiments of the present teachings;

[144] **FIG. 28G** shows fluorescent images two single cells with multiplexed gene tagging of ACTB (EGFP) and LMNB1 (mCherry) using PASTE according to embodiments of the present teachings;

[145] **FIG. 29A** shows the prime editing efficiency of Bxb1 attB site insertion at the ACTB locus according to embodiments of the present teachings;

[146] **FIG. 29B** shows the prime editing efficiency at inserting Bxb1 attB sites of different lengths at the ACTB locus according to embodiments of the present teachings;

[147] **FIG. 29C** shows the prime editing efficiency of inserting attB sequences from different integrases, wherein both orientations of landing sites are profiled (F, forward; and R, reverse) according to embodiments of the present teachings;

[148] **FIG. 29D** shows the prime editing efficiency of inserting attB sequences from Bxb1 integrase and Cre recombinase, wherein both orientations of landing sites are profiled (F, forward; and R, reverse) according to embodiments of the present teachings;

[149] **FIG. 29E** shows a schematic of PASTE insertion at the ACTB locus showing guide and target sequences according to embodiments of the present teachings. FIG. 29E discloses SEQ ID NOS 428-431, respectively, in order of appearance;

[150] **FIG. 29F** shows a comparison of PASTE integration efficiency of GFP with a panel of integrases targeting the 5' end of the ACTB locus, wherein both orientations of landing sites are profiled (F, forward; and R, reverse) according to embodiments of the present teachings;

[151] **FIG. 29G** shows a comparison of GFP cargo integration efficiency between Bxb1 integrases and Cre recombinase according to embodiments of the present teachings;

[152] **FIG. 29H** shows the dependence of PASTE editing activity on different prime and integrase components according to embodiments of the present teachings;

[153] **FIG. 29I** shows a titration of a single vector PASTE system (SpCas9-RT-P2A-Bxb1) on integrase efficiency according to embodiments of the present teachings;

[154] **FIG. 29J** shows the effect of cargo size on PASTE insertion efficiency at the endogenous ACTB target according to embodiments of the present teachings;

[155] **FIG. 29K** shows a gel electrophoresis showing complete insertion by PASTE for multiple cargo sizes according to embodiments of the present teachings;

[156] **FIG. 30A** shows a schematic of PASTE integration, including resulting attR and attL sites that are generated and PCR primers for assaying the integration junctions according to embodiments of the present teachings;

[157] **FIG. 30B** shows a PCR and gel electrophoresis readout of left integration junction from PASTE insertion of GFP at the ACTB locus, wherein the insertion is analyzed for in-frame and out-of-frame GFP integration experiments as well as for a no prime control and expected sizes of the PCR fragments are shown using the primers shown in the schematic in subpanel FIG. 30A according to embodiments of the present teachings;

[158] **FIG. 30C** shows a PCR and gel electrophoresis readout of right integration junction from PASTE insertion of GFP at the ACTB locus, wherein the insertion is analyzed for in-frame and out-of-frame GFP integration experiments as well as for a no prime control and the expected sizes of the PCR fragments are shown using the primers shown in the schematic in subpanel FIG. 30A according to embodiments of the present teachings;

[159] **FIG. 30D** shows a Sanger sequencing shown for the right integration junction for an in-frame fusion of GFP via PASTE to the N-terminus of ACTB according to embodiments of the present teachings;

[160] **FIG. 30E** shows a Sanger sequencing shown for the left integration junction for an in-frame fusion of GFP via PASTE to the N-terminus of ACTB according to embodiments of the present teachings;

[161] **FIG. 31A** shows a schematic of various parameters that affect PASTE integration of ~1 kb GFP insert, wherein on the pegRNA, the PBS, RT, and attB lengths can alter the efficiency of attB insertion, and nicking guide selection also affects overall gene integration efficiency according to embodiments of the present teachings;

[162] **FIG. 31B** shows the impact of PBS and RT length on PASTE integration of *GFP* at the *ACTB* locus according to embodiments of the present teachings;

[163] **FIG. 31C** shows the impact of PBS and RT length on PASTE integration of GFP at the *LMNB1* locus according to embodiments of the present teachings;

[164] **FIG. 31D** shows the impact of attB length on PASTE integration of *GFP* at the *ACTB* locus according to embodiments of the present teachings;

[165] **FIG. 31E** shows the impact of attB length on PASTE integration of GFP at the *LMNB1* locus according to embodiments of the present teachings;

[166] **FIG. 31F** shows the impact of attB length on PASTE integration of *GFP* at the *NOLC1* locus according to embodiments of the present teachings;

[167] **FIG. 31G** shows the impact of minimal PBS, RT, and attB lengths on PASTE integration efficiency of *GFP* at the *ACTB* locus according to embodiments of the present teachings;

[168] **FIG. 31H** shows the impact of minimal PBS, RT, and attB lengths on PASTE integration efficiency of *GFP* at the *LMNB1* locus according to embodiments of the present teachings;

[169] **FIG. 31I** shows the PASTE integration of *GFP* at the *LMNB1* locus in the presence and absence of nicking guide, prime, and Bxb1 with a minimally compact pegRNA containing a 38 bp attB compared to a longer pegRNA design according to embodiments of the present teachings;

[170] **FIG. 32A** shows the PASTE insertion efficiency at *ACTB* and *LMNB1* loci with two different nicking guide designs according to embodiments of the present teachings;

[171] **FIG. 32B** shows the PASTE editing efficiency at *ACTB* and *LMNB1* with target and non-targeting spacers and matched pegRNAs with and without Bxb1 expression according to embodiments of the present teachings;

[172] **FIG. 33A** shows the PASTE integration of GFP at the ACTB locus with different Bxb1 catalytic mutants according to embodiments of the present teachings;

[173] **FIG. 33B** shows the PASTE integration of GFP at the ACTB locus with different RT catalytic mutants according to embodiments of the present teachings;

[174] **FIG. 34A** shows the *GFP* integration by PASTE at a panel of endogenous genomic loci according to embodiments of the present teachings;

[175] **FIG. 34B** shows the integration of a panel of different gene cargo at *ACTB* locus via PASTE according to embodiments of the present teachings;

[176] **FIG. 34C** shows the integration efficiency of therapeutically relevant genes at the *ACTB* locus according to embodiments of the present teachings;

[177] **FIG. 34D** shows the endogenous protein tagging with *GFP* via PASTE by in-frame endogenous gene tagging at the *ACTB* loci and *SRRM2* loci according to embodiments of the present teachings;

[178] **FIG. 34E** shows the endogenous protein tagging with *GFP* via PASTE by in-frame endogenous gene tagging at the *NOLC1* loci and *LMNB1* loci according to embodiments of the present teachings;

[179] **FIG. 35** shows the integration of a panel of different gene cargo at *LMNB1* locus via PASTE according to embodiments of the present teachings;

[180] **FIG. 36A** shows the PASTE integration efficiency for all 16 central dinucleotide attB/attP sequence pairs with a 5 kb *GFP* template at the *ACTB* locus according to embodiments of the present teachings;

[181] **FIG. 36B** shows a schematic of the pooled attB/attP dinucleotide orthogonality assay, wherein each attB dinucleotide sequence is co-transfected with a barcoded pool of all 16 attP dinucleotide sequences and Bxb1 integrase, relative integration efficiencies are determined by next generation sequencing of barcodes, and all 16 attB dinucleotides are profiled in an arrayed format with attP pools according to embodiments of the present teachings;

[182] **FIG. 36C** shows the relative insertion preferences for all possible attB/attP dinucleotide pairs determined by the pooled orthogonality assay according to embodiments of the present teachings;

[183] **FIG. 36D** shows the orthogonality of top 4 attB/attP dinucleotide pairs evaluated for GFP integration with PASTE at the *ACTB* locus according to embodiments of the present teachings;

[184] **FIG. 37** shows the orthogonality of Bxb1 dinucleotides as measured by a pooled reporter assay, wherein each web logo motif shows the relative integration of different attP sequences in a pool at a denoted attB sequence with the listed dinucleotide according to embodiments of the present teachings;

[185] **FIG. 38A** shows a schematic of multiplexed integration of different cargo sets at specific genomic loci, wherein three fluorescent cargos (GFP, mCherry, and YFP) are inserted orthogonally at three different loci (ACTB, LMNB1, NOLC1) for in-frame gene tagging according to embodiments of the present teachings;

[186] **FIG. 38B** shows the efficiency of multiplexed PASTE insertion of combinations of fluorophores at *ACTB, LMNB1,* and *NOLC1* loci according to embodiments of the present teachings;

[187] **FIG. 39A** shows the *GFP* integration efficiency at a panel of genomic loci by PASTE compared to insertion rates by homology-independent targeted integration (HITI) according to embodiments of the present teachings;

[188] **FIG. 39B** shows a comparison of unintended indel generation by PASTE and HITI at the *ACTB* and *LMNB1* target sites, wherein the on-target EGFP integration rate observed compared to unintended indels is shown according to embodiments of the present teachings;

[189] **FIG. 39C** shows the integration of a *GFP* template by PASTE at the *ACTB* locus compared to homology-directed repair (HDR) at the same target, wherein the quantification is by single-cell clone counting, wherein targeting and non-targeting guides were used for HDR insertion, and wherein for PASTE targeting and non-targeting refers to the presence or absence of the SpCas9-RT protein respectively according to embodiments of the present teachings;

[190] **FIG. 39D** shows the comparison of unintended indel generation by PASTE and HDR based EGFP insertion at the *ACTB* target site, wherein the average indel rate measured across all single-cell clones generated is showed according to embodiments of the present teachings;

[191] **FIG. 39E** shows a schematic for Bxb1 and Cas9 off-target identification and a detection assay according to embodiments of the present teachings;

[192] **FIG. 39F** shows the GFP integration activity at predicted Bxb1 off-target sites in the human genome according to embodiments of the present teachings;

[193] **FIG. 39G** shows the GFP integrations activity at predicted PASTE *ACTB* Cas9 guide off target sites according to embodiments of the present teachings;

[194] **FIG. 39H** shows the GFP integration activity at predicted HITI *ACTB* Cas9 guide off-target sites according to embodiments of the present teachings;

[195] **FIG. 39I** shows a schematic of next-generation sequencing method to assay genome-wide off-target integration sites by PASTE according to embodiments of the present teachings;

[196] **FIG. 39J** shows the alignment of reads at the on-target *ACTB* site using a genome-wide integration assay, wherein expected on-target integration outcomes are shown according to embodiments of the present teachings;

[197] **FIG. 39K** shows the analysis of on-target and off-target integration events across 3 single-cell clones for PASTE and 3 single-cell clones for no prime condition according to embodiments of the present teachings;

[198] **FIG. 39L** shows a Manhattan plot of integration events for a representative single-cell clone with PASTE editing, wherein the on-target site is at the *ACTB* gene on chromosome 7 according to embodiments of the present teachings;

[199] **FIG. 40A** shows a comparison of indel rates generated by PASTE and HITI mediated insertion of EGFP at the *ACTB* and *LMNB1* loci in HepG2 cells according to embodiments of the present teachings;

[200] **FIG. 40B** shows the validation of ddPCR assays for detecting editing at predicted Bxb1 offtarget sites using synthetic amplicons according to embodiments of the present teachings;

[201] **FIG. 40C** shows the validation of ddPCR assays for detecting editing at predicted PASTE ACTB Cas9 guide off-target sites using synthetic amplicons according to embodiments of the present teachings;

[202] **FIG. 40D** shows the validation of ddPCR assays for detecting editing at predicted HITI ACTB Cas9 guide off-target sites using synthetic amplicons according to embodiments of the present teachings;

[203] **FIG. 41A** shows a number of significant differentially regulated genes in HEK293FT cells expressing Bxb1 integrase, PASTE targeting *ACTB* integration of EGFP, or Prime editing targeting *ACTB* for EGFP insertion without Bxb1 expression according to embodiments of the present teachings;

[204] **FIG. 41B** shows Volcano plots depicting the fold expression change of sequenced mRNAs versus significance (p-value), wherein each dot represents a unique mRNA transcript and significant transcripts are shaded according to either upregulation (red) or downregulation (blue), and wherein fold expression change is measured against *ACTB*-targeting guide-only expression (including cargo) according to embodiments of the present teachings;

[205] **FIG. 41C** shows top significantly upregulated and downregulated genes for Bxb1-only conditions, wherein genes are shown with their corresponding Z-scores of counts per million (cpm) for Bxb1 only expression, GFP-only expression, PASTE targeting ACTB for EGFP insertion, Prime targeting ACTB for EGFP expression without Bxb1, and guide/cargo only according to embodiments of the present teachings;

[206] **FIG. 42A** shows a schematic of PASTE performance in the presence of cell cycle inhibition, wherein cells are transfected with plasmids for insertion with PASTE or Cas9-induced HDR and treated with aphidicolin to arrest cell division, and wherein the efficiency

of PASTE and HDR are read out with ddPCR or amplicon sequencing respectively according to embodiments of the present teachings;

[207] FIG. 42B shows the editing efficiency of single mutations by HDR at EMX1 locus with two Cas9 guides in the presence or absence of cell division read out with amplicon sequencing according to embodiments of the present teachings;

[208] FIG. 42C shows the integration efficiency of various sized GFP inserts up to 13.3 kb at the *ACTB* locus with PASTE in the presence or absence of cell division according to embodiments of the present teachings;

[209] FIG. 42D shows the PASTE editing efficiency with two vector (PE2 and Bxb1) and single vector (PE2-P2A-Bxb1) designs in K562 cells according to embodiments of the present teachings;

[210] FIG. 42E shows the PASTE editing efficiency with single vector (PE2-P2A-Bxb1) designs in primary human T cells according to embodiments of the present teachings;

[211] FIG. 42F shows the integration efficiency of therapeutically relevant genes at the *ACTB* locus according to embodiments of the present teachings;

[212] FIG. 42G shows a schematic of protein production assay for PASTE-integrated transgene, wherein *SERPINA1* and *CPS1* transgenes are tagged with HIBIT luciferase for readout with both ddPCR and luminescence according to embodiments of the present teachings;

[213] FIG. 42H shows the integration efficiency of SERPINA1 and CPS1 transgenes in HEK293FT cells at the *ACTB* locus according to embodiments of the present teachings;

[214] FIG. 42I shows the integration efficiency of SERPINA1 and CPS1 transgenes in HepG2 cells at the *ACTB* locus according to embodiments of the present teachings;

[215] FIG. 42J shows the intracellular levels of SERPINA1-HIBIT and CPS1-HIBIT in HepG2 cells according to embodiments of the present teachings;

[216] FIG. 42K shows the secreted levels of SERPINA1-HIBIT and CPS1-HIBIT in HepG2 cells according to embodiments of the present teachings;

[217] FIG. 43A shows the HDR mediated editing of the EMX1 locus that is significantly diminished in non-dividing HEK293FT cells blocked by 5 μM aphidicolin treatment according to embodiments of the present teachings;

[218] FIG. 43B shows the effect of insert minicircle DNA amount on PASTE-mediated insertion at the *ACTB* locus in dividing and nondividing HEK293FT cells blocked by 5 μM aphidicolin treatment according to embodiments of the present teachings;

[219] **FIG. 43C** shows the PASTE integration of GFP at the ACTB locus with the GFP template delivered via AAV, showing dose dependence of integration efficiency according to embodiments of the present teachings;

[220] **FIG. 44A** shows the PASTE integration activity at three endogenous loci comparing the normal PASTE SV40 NLS to a c-Myc NLS/variable bi-partite SV40 NLS design according to embodiments of the present teachings;

[221] **FIG. 44B** shows the PASTE integration activity at the ACTB locus with different GFP minicircle template amounts comparing the normal PASTE SV40 NLS to a c-Myc NLS/variable bi-partite SV40 NLS design according to embodiments of the present teachings;

[222] **FIG. 45** shows the improvement of the PASTE editing activity using a puromycin growth selection marker according to embodiments of the present teachings;

[223] **FIG. 46A** shows the integration of SERPINA1 and CPS1 genes that are HIBIT tagged as measured by a protein expression luciferase assay according to embodiments of the present teachings;

[224] **FIG. 46B** shows the integration of SERPINA1 and CPS1 genes that are HIBIT tagged as measured by a protein expression luciferase assay normalized to a standardized HIBIT ladder, enabling accurate quantification of protein levels according to embodiments of the present teachings;

[225] **FIG. 47A** shows optimization of PASTE constructs with a panel of linkers and reverse transcriptase (RT) modifications for EGFP integration at the *ACTB* locus, according to embodiments of the present teachings;

[226] **FIG. 47B** shows the effect of cargo size on PASTE insertion efficiency at the endogenous *ACTB* target. Cargos were transfected with fixed molar amounts, according to embodiments of the present teachings;

[227] **FIG. 48A** shows prime editing efficiency for the insertion of different length BxbINT AttB sites at *ACTB*, according to embodiments of the present teachings;

[228] **FIG. 48B** shows prime editing efficiency for the insertion of a BxbINT AttB site at *ACTB* with targeting and non-targeting guides, according to embodiments of the present teachings;

[229] **FIG. 48C** shows prime editing efficiency for the insertion of different integrases' (Bxb1, Tp9, and Bt1) AttB sites at *ACTB*. Both orientations of landing sites are profiled (F, forward; R, reverse), according to embodiments of the present teachings;

[230] **FIG. 48D** shows PASTE editing efficiency for the insertion of EGFP at *ACTB* with and without a nicking guide, according to embodiments of the present teachings; and

[231] **FIG. 49A** shows optimization of PASTE editing by dosage titration and protein optimization. PASTE integration efficiency of EGFP at *ACTB* measured with different doses of a single-vector delivery of components.

[232] **FIG. 49B** PASTE integration efficiency of EGFP at *ACTB* measured with different ratios of a single-vector delivery of components to the EGFP template vector.

[233] **FIG. 49C** PASTE integration efficiency of EGFP at *ACTB* with different RT domain fusions.

[234] **FIG. 49D** PASTE integration efficiency of EGFP at *ACTB* with different RT domain fusions and linkers.

[235] **FIG. 49E** PASTE integration efficiency of EGFP at *ACTB* with mutant RT domains.

[236] **FIG. 49F** PASTE integration efficiency of EGFP at *ACTB* with mutated BxbINT domains.

[237] **FIG. 50A** Insertion templates delivered via AAV transduction. PASTE editing machinery was delivered via transfection, and templates were co-delivered via AAV dosing at levels indicated.

[238] **FIG. 50B** Schematic of AdV delivery of the complete PASTE system with three viral vectors.

[239] **FIG. 50C** Integration efficiency of AdV delivery of integrase, guides, and cargo in HEK293FT and HepG2 cells. BxbINT and guide RNAs or cargo were delivered either via plasmid transfection (Pl), AdV transduction (AdV), or omitted (-). SpCas9-RT was only delivered as plasmid or omitted.

[240] **FIG. 50D** AdV delivery of all PASTE components in HEK293FT and HepG2 cells.

[241] **FIG. 50E** Schematic of mRNA and synthetic guide delivery of PASTE components.

[242] **FIG. 50F** Delivery of PASTE system components with mRNA and synthetic guides, paired with either AdV or plasmid cargo.

[243] **FIG. 50G** Delivery of circular mRNA with synthetic guides and either AdV or plasmid cargo.

[244] **FIG. 50H** PASTE editing efficiency with single vector designs in primary human T cells.

[245] **FIG. 50I** PASTE editing efficiency with single vector designs in primary human hepatocytes.

[246] **FIG. 51A** PASTE editing efficiency at the LMNB1 locus with 130 bp and 385 bp deletions of the first exon of LMNB1 with combined insertion of an attB sequence.

[247] **FIG. 51B** PASTE editing efficiency with a 130 bp deletion of the first exon of LMNB1 with a combined insertion of a 967 bp cargo using the PASTE system.

DETAILED DESCRIPTION

[248] It will be appreciated that for clarity, the following discussion will describe various aspects of embodiments of the applicant's teachings. It should be noted that the specific embodiments are not intended as an exhaustive description or as a limitation to the broader aspects discussed herein. One aspect described in conjunction with a particular embodiment is not necessarily limited to that embodiment and can be practiced with any other embodiment(s). Reference throughout this specification to "one embodiment", "an embodiment," "an example embodiment," means that a particular feature, structure or characteristic described in connection with the embodiment is included in at least one embodiment of the present disclosure. Thus, appearances of the phrases "in one embodiment," "in an embodiment," or "an example embodiment" in various places throughout this specification are not necessarily all referring to the same embodiment, but may. Furthermore, the particular feature, structures or characteristics may be combined in any suitable manner, as would be apparent to a person skilled in the art from this disclosure, in one or more embodiments.

General Definitions

[249] Unless defined otherwise, technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this disclosure pertains. Definitions of common terms and techniques in molecular biology may be found in Molecular Cloning: A Laboratory Manual, 2nd edition (1989) (Sambrook, Fritsch, and Maniatis); Molecular Cloning: A Laboratory Manual, 4th edition (2012) (Green and Sambrook); Current Protocols in Molecular Biology (1987) (F.M. Ausubel *et al.* eds.); the series Methods in Enzymology (Academic Press, Inc.): PCR 2: A Practical Approach (1995) (M.J. MacPherson, B.D. Hames, and G.R. Taylor eds.): Antibodies, A Laboratory Manual (1988) (Harlow and Lane, eds.): Antibodies A Laboratory Manual, 2nd edition 2013 (E.A. Greenfield ed.); Animal Cell Culture (1987) (R.I. Freshney, ed.); Benjamin Lewin, Genes IX, published by Jones and Bartlet, 2008 (ISBN 0763752223); Kendrew *et al.* (eds.), The

Encyclopedia of Molecular Biology, published by Blackwell Science Ltd., 1994 (ISBN 0632021829); Robert A. Meyers (ed.), Molecular Biology and Biotechnology: a Comprehensive Desk Reference, published by VCH Publishers, Inc., 1995 (ISBN 9780471185710); Singleton *et al.*, Dictionary of Microbiology and Molecular Biology 2nd ed., J. Wiley & Sons (New York, N.Y. 1994), March, Advanced Organic Chemistry Reactions, Mechanisms and Structure 4th ed., John Wiley & Sons (New York, N.Y. 1992); and Marten H. Hofker and Jan van Deursen, Transgenic Mouse Methods and Protocols, 2nd edition (2011).

[250] As used herein, the singular forms "a", "an," and "the" include both singular and plural referents unless the context clearly dictates otherwise. Thus, for example, reference to "a cell" includes a plurality of such cells.

[251] As used herein, the term "optional" or "optionally" means that the subsequent described event, circumstance or substituent may or may not occur, and that the description includes instances where the event or circumstance occurs and instances where it does not.

[252] The recitation of numerical ranges by endpoints includes all numbers and fractions subsumed within the respective ranges, as well as the recited endpoints.

[253] As used herein, the term "about" or "approximately" refers to a measurable value such as a parameter, an amount, a temporal duration, and the like, are meant to encompass variations of and from the specified value, such as variations of +/-10% or less, +/-5% or less, +/-1% or less, +/-0.5% or less, and +/-0.1% or less of and from the specified value, insofar such variations are appropriate to perform in the disclosure. It is to be understood that the value to which the modifier "about" or "approximately" refers is itself also specifically, and preferably, disclosed.

[254] It is noted that all publications and references cited herein are expressly incorporated herein by reference in their entirety. The publications discussed herein are provided solely for their disclosure prior to the filing date of the present application. Nothing herein is to be construed as an admission that the present disclosure is not entitled to antedate such publication. Further, the dates of publication provided may be different from the actual publication dates which may need to be independently confirmed.

Overview

[255] The embodiments disclosed herein provide non-naturally occurring or engineered systems, methods, and compositions for site-specific genetic engineering using

Programmable Addition via Site-Specific Targeting Elements (PASTE). A schematic diagram illustrating the concept of PASTE is shown in FIG. 1. As discussed in more details below, PASTE comprises the addition of an integration site into a target genome followed by the insertion of one or more genes of interest or one or more nucleic acid sequences of interest at the site. This process can be done as one or more reactions in a cell. The addition of the integration site into the target genome is done using gene editing technologies that include for example, without limitation, prime editing, recombinant adeno-associated virus (rAAV)-mediated nucleic acid integration, transcription activator-like effector nucleases (TALENS), and zinc finger nucleases (ZFNs). The integration of the transgene at the integration site is done using integrase technologies that include for example, without limitation, integrases, recombinases and reverse transcriptases. The necessary components for the site-specific genetic engineering disclosed herein comprise at least one or more nucleases, one or more gRNA, one or more integration enzymes, and one or more sequences that are complementary or associated to the integration site and linked to the one or more genes of interest or one or more nucleic acid sequences of interest to be inserted into the cell genome.

[256] An advantage of the non-naturally occurring or engineered systems, methods, and compositions for site-specific genetic engineering disclosed herein is programmable insertion of large elements without reliance on DNA damage responses.

[257] Another advantage of the non-naturally occurring or engineered systems, methods, and compositions for site-specific genetic engineering disclosed herein is facile multiplexing, enabling programmable insertion at multiple sites.

[258] Another advantage of the non-naturally occurring or engineered systems, methods, and compositions for site-specific genetic engineering disclosed herein is scalable production and delivery through minicircle templates.

Prime Editing

[259] The present disclosure provides non-naturally occurring or engineered systems, methods, and compositions for site-specific genetic engineering using gene editing technologies, such as prime editing, to add an integration site into a target genome. Prime editing will be discussed in more details below.

[260] Prime editing is a versatile and precise genome editing method that directly writes new genetic information into a specified DNA site. A schematic diagram illustrating the concept of prime editing is shown in FIG. 2. See, Anzalone, A.V., *et al.* "Search-and-replace

genome editing without double-strand breaks or donor DNA," *Nature* 576, 149–157 (2019). Prime editing uses a catalytically-impaired Cas9 endonuclease that is fused to an engineered reverse transcriptase (RT) and programmed with a prime-editing guide RNA (pegRNA). The skilled person in the art would appreciate that the pegRNA both specifies the target site and encodes the desired edit. The catalytically-impaired Cas9 endonuclease also comprises a Cas9 nickase that is fused to the reverse transcriptase. During genetic editing, the Cas9 nickase part of the protein is guided to the DNA target site by the pegRNA. The reverse transcriptase domain then uses the pegRNA to template reverse transcription of the desired edit, directly polymerizing DNA onto the nicked target DNA strand. The edited DNA strand replaces the original DNA strand, creating a heteroduplex containing one edited strand and one unedited strand. Afterward, the prime editor (PE) guides resolution of the heteroduplex to favor copying the edit onto the unedited strand, completing the process.

[261] The prime editors refer to a Moloney Murine Leukemia Virus (M-MLV) reverse transcriptase (RT) fused to a Cas9 H840A nickase. Fusing the RT to the C-terminus of the Cas9 nickase may result in higher editing efficiency. Such a complex is called PE1. The Cas9(H840A) can also be linked to a non-M-MLV reverse transcriptase such as a AMV-RT or XRT (Cas9(H840A)-AMV-RT or XRT). In some embodiments, Cas 9(H840A) can be replaced with Cas12a/b or Cas9(D10A). A Cas9 (wild type), Cas9(H840A), Cas9(D10A) or Cas 12a/b nickase fused to a pentamutant of M-MLV RT (D200N/ L603W/ T330P/ T306K/ W313F), having up to about 45-fold higher efficiency is called PE2. In some embodiments, the M-MLV RT comprise one or more of the mutations: Y8H, P51L, S56A, S67R, E69K, V129P, L139P, T197A, H204R, V223H, T246E, N249D, E286R, Q291I, E302K, E302R, F309N, M320L, P330E, L435G, L435R, N454K, D524A, D524G, D524N, E562Q, D583N, H594Q, E607K, D653N, and L671P. In some embodiments, the reverse transcriptase can also be a wild-type or modified transcription xenopolymerase (RTX), avian myeloblastosis virus reverse transcriptase (AMV-RT), Feline Immunodeficiency Virus reverse transcriptase (FIV-RT), FeLV-RT (Feline leukemia virus reverse transcriptase), HIV-RT (Human Immunodeficiency Virus reverse transcriptase), or Eubacterium rectale maturase RT (MarathonRT). PE3 involves nicking the non-edited strand, potentially causing the cell to remake that strand using the edited strand as the template to induce HR. The nicking of the non-edited strand can involve the use of a nicking guide RNA (ngRNA).

[262] Nicking the non-edited strand can increase editing efficiency. For example, nicking the non-edited strand can increase editing efficiency by about 1.1 fold, about 1.3 fold, about

1.5 fold, about 1.7 fold, about 1.9 fold, about 2.1 fold, about 2.3 fold, about 2.5 fold, about 2.7 fold, about 2.9 fold, about 3.1 fold, about 3.3 fold, about 3.5 fold, about 3.7 fold, about 3.9 fold, 4.1 fold, about 4.3 fold, about 4.5 fold, about 4.7 fold, about 4.9 fold, or any range that is formed from any two of those values as endpoints.

[263] Although the optimal nicking position varies depending on the genomic site, nicks positioned 3′ of the edit about 40–90 bp from the pegRNA-induced nick can generally increase editing efficiency without excess indel formation. The prime editing practice allows starting with non-edited strand nicks about 50 bp from the pegRNA-mediated nick, and testing alternative nick locations if indel frequencies exceed acceptable levels.

[264] As used herein, the term "guide RNA" (gRNA) and the like refer to a RNA that guide the insertion or deletion of one or more genes of interest or one or more nucleic acid sequences of interest into a target genome. The gRNA can also refer to a prime editing guide RNA (pegRNA), a nicking guide RNA (ngRNA), and a single guide RNA (sgRNA). In some embodiments, the term "gRNA molecule" refers to a nucleic acid encoding a gRNA. In some embodiments, the gRNA molecule is naturally occurring. In some embodiments, a gRNA molecule is non-naturally occurring. In some embodiments, a gRNA molecule is a synthetic gRNA molecule. A gRNA can target a nuclease or a nickase such as Cas9, Cas 12a/b, Cas9 (H840A) or Cas9 (D10A) molecule to a target nucleic acid or sequence in a genome. In some embodiments, the gRNA can bind to a DNA nickase bound to a reverse transcriptase domain. A "modified gRNA," as used herein, refers to a gRNA molecule that has an improved half-life after being introduced into a cell as compared to a non-modified gRNA molecule after being introduced into a cell. In some embodiments, the guide RNA can facilitate the addition of the insertion site sequence for recognition by integrases, transposases, or recombinases.

[265] As used herein, the term "prime-editing guide RNA" (pegRNA) and the like refer to an extended single guide RNA (sgRNA) comprising a primer binding site (PBS), a reverse transcriptase (RT) template sequence, and an integration site sequence that can be recognized by recombinases, integrases, or transposases. Exemplary design parameters for pegRNA are shown in FIG. 24A. For example, the PBS can have a length of at least about 4 nt, 5 nt, 6 nt, 7 nt, 8 nt, 9 nt, 10 nt, 11 nt, 12 nt, 13 nt, 14 nt, 15 nt, 16 nt, 17 nt, 18 nt, 19 nt, 20 nt, 21 nt, 22 nt, 23 nt, 24 nt, 25 nt, 26 nt, 27 nt, 28 nt, 29 nt, 30 nt, or more nt. For example, the PBS can have a length of about 4 nt, 5 nt, 6 nt, 7 nt, 8 nt, 9 nt, 10 nt, 11 nt, 12 nt, 13 nt, 14 nt, 15 nt, 16 nt, 17 nt, 18 nt, 19 nt, 20 nt, 21 nt, 22 nt, 23 nt, 24 nt, 25 nt, 26 nt, 27 nt, 28 nt, 29 nt, 30 nt, or any range that is formed from any two of those values as endpoints. For example, the RT

template sequence can have a length of at least about 4 nt, 5 nt, 6 nt, 7 nt, 8 nt, 9 nt, 10 nt, 11 nt, 12 nt, 13 nt, 14 nt, 15 nt, 16 nt, 17 nt, 18 nt, 19 nt, 20 nt, 21 nt, 22 nt, 23 nt, 24 nt, 25 nt, 26 nt, 27 nt, 28 nt, 29 nt, 30 nt, 31 nt, 32 nt, 33 nt, 34 nt, 35 nt, 36 nt, 37 nt, 38 nt, 39 nt, 40 nt, 41 nt, 42 nt, 43 nt, 44 nt, 45 nt, 46 nt, 47 nt, 48 nt, 49 nt, 50 nt, or more nt. For example, the RT template sequence can have a length of about 4 nt, 5 nt, 6 nt, 7 nt, 8 nt, 9 nt, 10 nt, 11 nt, 12 nt, 13 nt, 14 nt, 15 nt, 16 nt, 17 nt, 18 nt, 19 nt, 20 nt, 21 nt, 22 nt, 23 nt, 24 nt, 25 nt, 26 nt, 27 nt, 28 nt, 29 nt, 30 nt, 31 nt, 32 nt, 33 nt, 34 nt, 35 nt, 36 nt, 37 nt, 38 nt, 39 nt, 40 nt, 41 nt, 42 nt, 43 nt, 44 nt, 45 nt, 46 nt, 47 nt, 48 nt, 49 nt, 50 nt, or any range that is formed from any two of those values as endpoints.

[266] During genome editing, the primer binding site allows the 3' end of the nicked DNA strand to hybridize to the pegRNA, while the RT template serves as a template for the synthesis of edited genetic information. The pegRNA is capable for instance, without limitation, of (i) identifying the target nucleotide sequence to be edited and (ii) encoding new genetic information that replaces the targeted sequence. In some embodiments, the pegRNA is capable of (i) identifying the target nucleotide sequence to be edited and (ii) encoding an integration site that replaces the targeted sequence.

[267] As used herein, the term "nicking guide RNA" (ngRNA) and the like refer to an RNA sequence that can nick a strand such as an edited strand and a non-edited strand. Exemplary design parameters for ngRNA are shown in FIG. 24B. The ngRNA can induce nicks at about 1 or more nt away from the site of the gRNA-induced nick. For example, the ngRNA can nick at least at about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, or more nt away from the site of the gRNA induced nick. In some embodiments, the ngRNA comprises SEQ ID NO: 75 with guide sequence SEQ ID NO: 74. As used herein, the terms "reverse transcriptase" and "reverse transcriptase domain" refer to an enzyme or an enzymatically active domain that can reverse a RNA transcribe into a complementary DNA. The reverse transcriptase or reverse transcriptase domain is a RNA dependent DNA polymerase. Such reverse transcriptase domains encompass, but are not limited, to a M-MLV reverse transcriptase, or a modified reverse transcriptase such as, without limitation, Superscript® reverse transcriptase (Invitrogen; Carlsbad, California),

Superscript® VILO™ cDNA synthesis (Invitrogen; Carlsbad, California), RTX, AMV-RT, and Quantiscript Reverse Transcriptase (Qiagen, Hilden, Germany).

[268] The pegRNA-PE complex disclosed herein recognizes the target site in the genome and the Cas9 for example nicks a protospacer adjacent motif (PAM) strand. The primer binding site (PBS) in the pegRNA hybridizes to the PAM strand. The RT template operably linked to the PBS, containing the edit sequence, directs the reverse transcription of the RT template to DNA into the target site. Equilibration between the edited 3′ flap and the unedited 5′ flap, cellular 5′ flap cleavage and ligation, and DNA repair results in stably edited DNA. To optimize base editing, a Cas9 nickase can be used to nick the non-edited strand, thereby directing DNA repair to that strand, using the edited strand as a template.

Integrase Technologies

[269] The present disclosure provides non-naturally occurring or engineered systems, methods, and compositions for site-specific genetic engineering using integrase technologies. Integrase technologies will be discussed in more details below.

[270] The integrase technologies used herein comprise proteins or nucleic acids encoding the proteins that direct integration of a gene of interest or nucleic acid sequence of interest into an integration site via a nuclease such as a prime editing nuclease. The protein directing the integration can be an enzyme such as integration enzyme. The integration enzyme can be an integrase that incorporates the genome or nucleic acid of interest into the cell genome at the integration site by integration. The integration enzyme can be a recombinase that incorporates the genome or nucleic acid of interest into the cell genome at the integration site by recombination. The integration enzyme can be a reverse transcriptase that incorporates the genome or nucleic acid of interest into the cell genome at the integration site by reverse transcription. The integration enzyme can be a retrotransposase that incorporates the genome or nucleic acid of interest into the cell genome at the integration site by retrotransposition.

[271] As used herein, the term "integration enzyme" refers to an enzyme or protein used to integrate a gene of interest or nucleic acid sequence of interest into a desired location or at the integration site, in the genome of a cell, in a single reaction or multiple reactions. Example of integration enzymes include for example, without limitation, Cre, Dre, Vika, Bxb1, φC31, RDF, FLP, φBT1, R1, R2, R3, R4, R5, TP901-1, A118, φFC1, φC1, MR11, TG1, φ370.1, Wβ, BL3, SPBc, K38, Peaches, Veracruz, Rebeuca, Theia, Benedict, KSSJEB, PattyP, Doom, Scowl, Lockley, Switzer, Bob3, Troube, Abrogate, Anglerfish, Sarfire, SkiPole,

ConceptII, Museum, Severus, Airmid, Benedict, Hinder, ICleared, Sheen, Mundrea, BxZ2, φRV, and retrotransposases encoded by R2, L1, Tol2 Tc1, Tc3, Mariner (Himar 1), Mariner (mos 1), and Minos. In some embodiments, the term "integration enzyme" refers to a nucleic acid (DNA or RNA) encoding the above-mentioned enzymes. In some embodiments, the Cre recombinase is expressed from a Cre recombinase expression plasmid (SEQ ID NO: 71).

[272] Mammalian expression plasmids can be found in Table 1 below.

Table 1

Name	**Full Description**	**SEQ ID NOS:**
PE2-Bxb1 Single Vector	pCMV-PE2-P2A- Bxb1	(SEQ ID NO: 381)
PE2 prime editor	pCMV-PE2 / Addgene #132775	(SEQ ID NO: 382)
PE2*-Bxb1 Single Vector	New NLS pCMV- PE2-P2A-Bxb1	(SEQ ID NO: 383)
PASTEv3	pCMV-SpCas9-XTEN-RT(1-478)-Sto7d-GGGGS-BxbINT	(SEQ ID NO: 384)
ACTB pegRNA	ACTB N-term PBS 13 RT 29 attB 46 pegRNA	(SEQ ID NO: 385)
ACTB Nicking +48	ACTB N-term Nicking guide 1 +48 guide	(SEQ ID NO: 386)
Bxb1 integrase	pCAG-NLS-HA-Bxb1integrase / Addgene #51271	(SEQ ID NO: 387)
TP901-1 Integrase	TP901-1 Integrase	(SEQ ID NO: 388)

PhiBT Integrase	PhiBT Integrase	(SEQ ID NO: 389)
HDR sgRNA guide	Minicircle U6-sgRNA EFS-SpCas9	(SEQ ID NO: 390)
HDR EGFP cargo	Cas9 HDR template site with EGFP	(SEQ ID NO: 391)
AAV helper plasmid	PDF6 AAV helper plasmid	(SEQ ID NO: 392)
AAV EGFP donor	GFP AAV donor plasmid	(SEQ ID NO: 393)
AAV2/8	AAV2/8 capsid protein	(SEQ ID NO: 394)

[273] Minicircle cargo gene maps can be found in Table 2 below.

Table 2

Name	**Full Description**	**SEQ ID NOS:**
Cargo EGFP	Parent minicircle plasmid - Cargo EGFP with attP Bxb1 site	(SEQ ID NO: 76)
Cargo EGFP post cleavage	Cargo EGFP with attP Bxb1 site - post minicircle cleavage	(SEQ ID NO: 395)
Cargo EGFP for fusion	Parent minicircle plasmid - Cargo EGFP with attP Bxb1 site for fusion	(SEQ ID NO: 396)

mCherry Cargo post cleavage	Cargo mCherry with attP Bxb1 site - post minicircle cleavage	(SEQ ID NO: 397)
YFP Cargo post cleavage	Cargo YFP with attP Bxb1 site - post minicircle cleavage	(SEQ ID NO: 398)
SERPINA1 Cargo post cleavage	Cargo SERPINA1 with attP Bxb1 site - post minicircle cleavage	(SEQ ID NO: 399)
CPS1 Cargo post cleavage	Cargo CPS1 with attP Bxb1 site - post minicircle cleavage	(SEQ ID NO: 400)
CFTR Cargo	Parent minicircle plasmid - Cargo CFTR with attP Bxb1 site	(SEQ ID NO: 401)
NYESO TCR Cargo post cleavage	Cargo NYESO TCR with attP Bxb1 site - post minicircle cleavage	(SEQ ID NO: 402)

[274] In some embodiments, the serine integrase φC31 from φC31 phage is use as integration enzyme. The integrase φC31 in combination with a pegRNA can be used to insert the pseudo attP integration site (SEQ ID NO: 78). A DNA minicircle containing a gene or nucleic acid of interest and attB (SEQ ID NO: 3) site can be used to integrate the gene or

nucleic acid of interest into the genome of a cell. This integration can be aided by a co-transfection of an expression vector having the φC31 integrase.

[275] As used herein, the term "integrase" refers to a bacteriophage derived integrase, including wild-type integrase and any of a variety of mutant or modified integrases. As used herein, the term "integrase complex" may refer to a complex comprising integrase and integration host factor (IF). As used herein, the term "integrase complex" and the like may also refer to a complex comprising an integrase, an integration host factor, and a bacteriophage X-derived excisionase (Xis).

[276] As used herein, the term "recombinase" and the like refer to a site-specific enzyme that mediates the recombination of DNA between recombinase recognition sequences, which results in the excision, integration, inversion, or exchange (*e.g.*, translocation) of DNA fragments between the recombinase recognition sequences. Recombinases can be classified into two distinct families: serine recombinases (*e.g.*, resolvases and invertases) and tyrosine recombinases (*e.g.*, integrases). Examples of serine recombinases include, without limitation, Hin, Gin, Tn3, β-six, CinH, ParA, γδ, Bxb1, φC31, TP901, TG1, φBT1, R1, R2, R3, R4, R5, φRV1, φFC1, MR11, A118, U153, and gp29. Examples of serine recombinases also include, without limitation, recombinases Peaches, Veracruz, Rebeuca, Theia, Benedict, KSSJEB, PattyP, Doom, Scowl, Lockley, Switzer, Bob3, Troube, Abrogate, Anglerfish, Sarfire, SkiPole, ConceptII, Museum, Severus, Airmid, Benedict, Hinder, ICleared, Sheen, Mundrea, and BxZ2 from Mycobacterial phages. Examples of tyrosine recombinases include, without limitation, Cre, FLP, R, Lambda, HK101, HK022, and pSAM2. The serine and tyrosine recombinase names stem from the conserved nucleophilic amino acid residue that the recombinase uses to attack the DNA and which becomes covalently linked to the DNA during strand exchange.

[277] Recombinases have numerous applications, including the creation of gene knockouts/knock-ins and gene therapy applications. See, *e.g.*, Brown *et al.*, "Serine recombinases as tools for genome engineering." *Methods*, 2011; 53(4):372-9; Hirano *et al.*, "Site-specific recombinases as tools for heterologous gene integration." Appl. Microbiol. Biotechnol. 2011; 92(2):227-39; Chavez and Calos, "Therapeutic applications of the ΦC31 integrase system." *Curr. Gene Ther.* 2011; 11(5):375-81; Turan and Bode, "Site-specific recombinases: from tag-and-target- to tag-and-exchange-based genomic modifications." *FASEB J.* 2011; 25(12):4088-107; Venken and Bellen, "Genome-wide manipulations of Drosophila melanogaster with transposons, Flp recombinase, and ΦC31 integrase." *Methods*

Mol. Biol. 2012; 859:203-28; Murphy, "Phage recombinases and their applications." *Adv. Virus Res.* 2012; 83:367-414; Zhang *et al.*, "Conditional gene manipulation: Creating a new biological era." J. Zhejiang Univ. Sci. B. 2012; 13(7):511-24; Karpenshif and Bernstein, "From yeast to mammals: recent advances in genetic control of homologous recombination." DNA Repair (Amst). 2012; 1; 11(10):781-8; the entire contents of each are hereby incorporated by reference in their entirety.

[278] The recombinases provided herein are not meant to be exclusive examples of recombinases that can be used in embodiments of the disclosure. The methods and compositions of the disclosure can be expanded by mining databases for new orthogonal recombinases or designing synthetic recombinases with defined DNA specificities (See, *e.g.*, Groth *et al.*, "Phage integrases: biology and applications." *J. Mol. Biol.* 2004; 335, 667-678; Gordley *et al.*, "Synthesis of programmable integrases." *Proc. Natl. Acad. Sci.* USA. 2009; 106, 5053-5058; the entire contents of each are hereby incorporated by reference in their entirety).

[279] Other examples of recombinases that are useful in the systems, methods, and compositions described herein are known to those of skill in the art, and any new recombinase that is discovered or generated is expected to be able to be used in the different embodiments of the disclosure.

[280] As used herein, the term "retrotransposase" and the like refer to an enzyme, or combination of one or more enzymes, wherein at least one enzyme has a reverse transcriptase domain. Retrotransposases are capable of inserting long sequences (e.g., over 3000 nucleotides) of heterologous nucleic acid into a genome. Examples of retrotransposases include for example, without limitation, retrotransposases encoded by elements such as R2, L1, Tol2 Tc1, Tc3, Mariner (Himar 1), Mariner (mos 1), Minos, and any mutants thereof.

[281] In some embodiments, the one or more genes of interest or one or more nucleic acid sequences of interest are inserted into a desired location in a genome using a RNA fragment, such as a retrotransposon, encoding the nucleic acid linked to a complementary or associated integration site. The insertion of the nucleic acid of interest into a location in the desired location in the genome using a retrotransposon is aided by a retrotransposase.

[282] The gene and nucleic acid sequence of interest disclosed herein can be any gene and nucleic acid sequence that are known in the art. The gene and nucleic acid sequence of interest can be for therapeutic and/or diagnostic uses. Examples of genes of interest include, without limitation, GBA, BTK, ADA, CNGB3, CNGA3, ATF6, GNAT2, ABCA1, ABCA7,

APOE, CETP, LIPC, MMP9, PLTP, VTN, ABCA4, MFSD8, TLR3, TLR4, ERCC6, HMCN1, HTRA1, MCDR4, MCDR5, ARMS2, C2, C3, CFB, CFH, JAG1, NOTCH2, CACNA1F, SERPINA1, TTR, GSN, B2M, APOA2, APOA1, OSMR, ELP4, PAX6, ARG, ASL, PITX2, FOXC1, BBS1, BBS10, BBS2, BBS9, MKKS, MKS1, BBS4, BBS7, TTC8, ARL6, BBS5, BBS12, TRIM32, CEP290, ADIPOR1, BBIP1, CEP19, IFT27, LZTFL1, DMD, BEST1, HBB, CYP4V2, AMACR, CYP7B1, HSD3B7, AKR1D1, OPN1SW, NR2F1, RLBP1, RGS9, RGS9BP, PROM1, PRPH2, GUCY2D, CACD, CHM, ALAD, ASS1, SLC25A13, OTC, ACADVL, ETFDH, TMEM67, CC2D2A, RPGRIP1L, KCNV2, CRX, GUCA1A, CERKL, CDHR1, PDE6C, TTLL5, RPGR, CEP78, C21orf2, C8ORF37, RPGRIP1, ADAM9, POC1B, PITPNM3, RAB28, CACNA2D4, AIPL1, UNC119, PDE6H, OPN1LW, RIMS1, CNNM4, IFT81, RAX2, RDH5, SEMA4A, CORD17, PDE6B, GRK1, SAG, RHO, CABP4, GNB3, SLC24A1, GNAT1, GRM6, TRPM1, LRIT3, TGFBI, TACSTD2, KRT12, OVOL2, CPS1, UGT1A1, UGT1A9, UGT1A8, UGT1A7, UGT1A6, UGT1A5, UGT1A4, CFTR, DLD, EFEMP1, ABCC2, ZNF408, LRP5, FZD4, TSPAN12, EVR3, APOB, SLC2A2, LOC106627981, GBA1, NR2E3, OAT, SLC40A1, F8, F9, UROD, CPOX, HFE, JH, LDLR, EPHX1, TJP2, BAAT, NBAS, LARS1, HAMP, HJV, RS1, ADAMTS18, LRAT, RPE65, LCA5, MERTK, GDF6, RD3, CCT2, CLUAP1, DTHD1, NMNAT1, SPATA7, IFT140, IMPDH1, OTX2, RDH12, TULP1, CRB1, MT-ND4, MT-ND1, MT-ND6, BCKDHA, BCKDHB, DBT, MMAB, ARSB, GUSB, NAGS, NPC1, NPC2, NDP, OPA1, OPA3, OPA4, OPA5, RTN4IP1, TMEM126A, OPA6, OPA8, ACO2, PAH, PRKCSH, SEC63, GAA, UROS, PPOX, HPX, HMOX1, HMBS, MIR223, CYP1B1, LTBP2, AGXT, ATP8B1, ABCB11, ABCB4, FECH, ALAS2, PRPF31, RP1, EYS, TOPORS, USH2A, CNGA1, C2ORF71, RP2, KLHL7, ORF1, RP6, RP24, RP34, ROM1, ADGRA3, AGBL5, AHR, ARHGEF18, CA4, CLCC1, DHDDS, EMC1, FAM161A, HGSNAT, HK1, IDH3B, KIAA1549, KIZ, MAK, NEUROD1, NRL, PDE6A, PDE6G, PRCD, PRPF3, PRPF4, PRPF6, PRPF8, RBP3, REEP6, SAMD11, SLC7A14, SNRNP200, SPP2, ZNF513, NEK2, NEK4, NXNL1, OFD1, RP1L1, RP22, RP29, RP32, RP63, RP9, RGR, POMGNT1, DHX38, ARL3, COL2A1, SLCO1B1, SLCO1B3, KCNJ13, TIMP3, ELOVL4, TFR2, FAH, HPD, MYO7A, CDH23, PCDH15, DFNB31, GPR98, USH1C, USH1G, CIB2, CLRN1, HARS, ABHD12, ADGRV1, ARSG, CEP250, IMPG1, IMPG2, VCAN, G6PC1, ATP7B and any derivatives thereof.

[283] As used here, the terms "retrotransposons," "jumping genes," "jumping nucleic acids," and the like refer to cellular movable genetic elements dependent on reverse

transcription. The retrotransposons are of non-replication competent cellular origin, and are capable of carrying a foreign nucleic acid sequence. The retrotransposons can act as parasites of retroviruses, retaining certain classical hallmarks, such as long terminal repeats (LTR), retroviral primer binding sites, and the like. However, the naturally occurring retrotransposons usually do not contain functional retroviral structure genes, which would normally be capable of recombining to yield replication competent viruses. Some retrotransposons are examples of so-called "selfish DNA", or genetic information, which encodes nothing except the ability to replicate itself. The retrotransposon may do so by utilizing the occasional presence of a retrovirus or a retrotransposase within the host cell, efficiently packaging itself within the viral particle, which transports it to the new host genome, where it is expressed again as RNA. The information encoded within that RNA is potentially transported with the jumping gene. A retrotransposon can be a DNA transposon or a retrotransposon, including a LTR retrotransposon or a non-LTR retrotransposon.

[284] Non-long terminal repeat (LTR) retrotransposons are a type of mobile genetic elements that are widespread in eukaryotic genomes. They include two classes: the apurinic/apyrimidinic endonuclease (APE)-type and the restriction enzyme-like endonuclease (RLE)-type. The APE class retrotransposons are comprised of two functional domains: an endonuclease/DNA binding domain, and a reverse transcriptase domain. The RLE class are comprised of three functional domains: a DNA binding domain, a reverse transcription domain, and an endonuclease domain. The reverse transcriptase domain of non-LTR retrotransposon functions by binding an RNA sequence template and reverse transcribing it into the host genome's target DNA. The RNA sequence template has a 3′ untranslated region which is specifically bound to the transposase, and a variable 5′ region generally having Open Reading Frame(s) ("ORF") encoding transposase proteins. The RNA sequence template may also comprise a 5′ untranslated region which specifically binds the retrotransposase. In some embodiments, a non-LTR transposons can include a LINE retrotransposon, such as L1, and a SINE retrotransposon, such as an Alu sequence. Other examples include for example, without limitation, R1, R2, R3, R4, and R5 retro-transposons (Moss, W. N. *et al., RNA Biol.* 2011, 8(5), 714–718; and Burke, W. D. *et al., Molecular Biology and Evolution* 2003, 20(8), 1260–1270). The transposon can be autonomous or non-autonomous.

[285] LTR retrotransposons, which include retroviruses, make up a significant fraction of the typical mammalian genome, comprising about 8% of the human genome and 10% of the

mouse genome. Lander *et al.*, 2001, *Nature* 409, 860-921; Waterson et al., 2002, *Nature* 420, 520-562. LTR elements include retrotransposons, endogenous retroviruses (ERVs), and repeat elements with HERV origins, such as SINE-R. LTR retrotransposons include two LTR sequences that flank a region encoding two enzymes: integrase and retrotransposase.

[286] ERVs include human endogenous retroviruses (HERVs), the remnants of ancient germ-cell infections. While most HERV proviruses have undergone extensive deletions and mutations, some have retained ORFS coding for functional proteins, including the glycosylated env protein. The env gene confers the potential for LTR elements to spread between cells and individuals. Indeed, all three open reading frames (pol, gag, and env) have been identified in humans, and evidence suggests that ERVs are active in the germline. See, e.g., Wang *et al.*, 2010, *Genome Res.* 20, 19-27. Moreover, a few families, including the HERV-K (HML-2) group, have been shown to form viral particles, and an apparently intact provirus has recently been discovered in a small fraction of the human population. See, e.g., Bannert and Kurth, 2006, Proc. Natl. Acad. USA 101, 14572-14579.

[287] LTR retrotransposons insert into new sites in the genome using the same steps of DNA cleavage and DNA strand-transfer observed in DNA transposons. In contrast to DNA transposons, however, recombination of LTR retrotransposons involves an RNA intermediate. LTR retrotransposons make up about 8% of the human genome. See, e.g., Lander et al., 2001, *Nature* 409, 860-921; Hua-Van et al., 2011, Biol. Dir. 6, 19.

Integration Site

[288] The present disclosure provides non-naturally occurring or engineered systems, methods, and compositions for site-specific genetic engineering via the addition of an integration site into a target genome. The integration site will be discussed in more details below.

[289] As used herein, the term "integration site" refers to the site within the target genome where one or more genes of interest or one or more nucleic acid sequences of interest are inserted. Examples of integration sites include for example, without limitation, a lox71 site (SEQ ID NO: 1), attB sites (SEQ ID NO: 3 and SEQ ID NO: 43), attP sites (SEQ ID NO: 4 and SEQ ID NO: 44), an attL site (SEQ ID NO: 67), an attR site (SEQ ID NO: 68), a Vox site (SEQ ID NO: 69), a FRT site (SEQ ID NO: 70), or a pseudo attP site (SEQ ID NO: 78). The integration site can be inserted into the genome or a fragment thereof of a cell using a

nuclease, a gRNA, and/or an integration enzyme. The integration site can be inserted into the genome of a cell using a prime editor such as, without limitation, PE1, PE2, and PE3, wherein the integration site is carried on a pegRNA. The pegRNA can target any site that is known in the art. Examples of cites targeted by the pegRNA include, without limitation, ACTB, SUPT16H, SRRM2, NOLC1, DEPDC4, NES, LMNB1, AAVS1 locus, CC10, CFTR, SERPINA1, ABCA4, and any derivatives thereof. The complementary integration site may be operably linked to a gene of interest or nucleic acid sequence of interest in an exogenous DNA or RNA. In some embodiments, one integration site is added to a target genome. In some embodiments, more than one integration sites are added to a target genome.

[290] To insert multiple genes or nucleic acids of interest, two or more integration sites are added to a desired location. Multiple DNA comprising nucleic acid sequences of interest are flanked orthogonal to the integration sequences, such as, without limitation, attB and attP. An integration site is "orthogonal" when it does not significantly recognize the recognition site or nucleotide sequence of a recombinase. Thus, one attB site of a recombinase can be orthogonal to an attB site of a different recombinase. In addition, one pair of attB and attP sites of a recombinase can be orthogonal to another pair of attB and attP sites recognized by the same recombinase. A pair of recombinases are considered orthogonal to each other, as defined herein, when there is recognition of each other' s attB or attP site sequences.

[291] The lack of recognition of integration sites or pairs of sites by the same recombinase or a different recombinase can be less than about 30%. In some embodiments, the lack of recognition of integration sites or pairs of sites by the same recombinase or a different recombinase can be less than about 30%, less than about 28%, less than about 26%, less than about 24%, less than about 22%, less than about 20%, less than about 18%, less than about 16%, less than about 14%, less than about 12%, less than about 10%, less than about 8%, less than about 6%, less than about 4%, less than about 2%, about 1%, or any range that is formed from any two of those values as endpoints. The crosstalk can be less than about 30%. In some embodiments, the crosstalk is less than about 30%, less than about 28%, less than about 26%, less than about 24%, less than about 22%, less than about 20%, less than about 18%, less than about 16%, less than about 14%, less than about 12%, less than about 10%, less than about 8%, less than about 6%, less than about 4%, less than about 2%, less than about 1%, or any range that is formed from any two of those values as endpoints.

[292] In some embodiments, the attB and/or attP site sequences comprise a central dinucleotide sequence. It has been shown that, for example, the central dinucleotide can be

changed to GA from GT and that only GA containing attB/attP sites interact and will not cross react with GT containing sequences. In some embodiments, the central dinucleotide is selected from the group consisting of AG, AC, TG, TC, CA, CT, GA, AA, TT, CC, GG, AT, TA, GC, CG and GT.

[293] As used herein, the term "pair of an attB and attP site sequences" and the like refer to attB and attP site sequences that share the same central dinucleotide and can recombine. This means that in the presence of one serine integrase as many as six pairs of these orthogonal att sites can recombine (attPTT will specifically recombine with attBTT, attPTC will specifically recombine with attBTC, and so on).

[294] In some embodiments, the central dinucleotide is nonpalindromic. In some embodiments, the central dinucleotide is palindromic. In some embodiments, a pair of an attB site sequence and an attP site sequence are used in different DNA encoding genes of interest or nucleic acid sequences of interest for inducing directional integration of two or more different nucleic acids.

[295] The Table 3 below shows examples of pairs of attB site sequence and attP site sequence with different central dinucleotide (CD).

Table 3

Pair	attB	attP	CD
1	SEQ ID NO: 5	SEQ ID NO: 6	TT
2	SEQ ID NO: 7	SEQ ID NO: 8	AA
3	SEQ ID NO: 9	SEQ ID NO: 10	CC
4	SEQ ID NO: 11	SEQ ID NO: 12	GG
5	SEQ ID NO: 13	SEQ ID NO: 14	TG
6	SEQ ID NO: 15	SEQ ID NO: 16	GT
7	SEQ ID NO: 17	SEQ ID NO: 18	CT
8	SEQ ID NO: 19	SEQ ID NO: 20	CA
9	SEQ ID NO: 21	SEQ ID NO: 22	TC

10	SEQ ID NO: 23	SEQ ID NO: 24	GA
11	SEQ ID NO: 25	SEQ ID NO: 26	AG
12	SEQ ID NO: 27	SEQ ID NO: 28	AC
13	SEQ ID NO: 29	SEQ ID NO: 30	AT
14	SEQ ID NO: 31	SEQ ID NO: 32	GC
15	SEQ ID NO: 33	SEQ ID NO: 34	CG
16	SEQ ID NO: 35	SEQ ID NO: 36	TA

PASTE

[296] The present disclosure provides non-naturally occurring or engineered systems, methods, and compositions for site-specific genetic engineering using PASTE. PASTE will be discussed in more details below.

[297] The site-specific genetic engineering disclosed herein is for the insertion of one or more genes of interest or one or more nucleic acid sequences of interest into a genome of a cell. In some embodiments, the gene of interest is a mutated gene implicated in a genetic disease such as, without limitation, a metabolic disease, cystic fibrosis, muscular dystrophy, hemochromatosis, Tay-Sachs, Huntington disease, Congenital Deafness, Sickle cell anemia, Familial hypercholesterolemia, adenosine deaminase (ADA) deficiency, X-linked SCID (X-SCID), and Wiskott-Aldrich syndrome (WAS). In some embodiments, the gene of interest or nucleic acid sequence of interest can be a reporter gene upstream or downstream of a gene for genetic analyses such as, without limitation, for determining the expression of a gene. In some embodiments, the reporter gene is a GFP template (SEQ ID NO: 76) or a Gaussia Luciferase (G-Luciferase) template (SEQ ID NO: 77) In some embodiments, the gene of interest or nucleic acid sequence of interest can be used in plant genetics to insert genes to enhance drought tolerance, weather hardiness, and increased yield and herbicide resistance in plants. In some embodiments, the gene of interest or nucleic acid sequence of interest can be used for site-specific insertion of a protein (*e.g.*, a lysosomal enzyme), a blood factor (*e.g.*, Factor I, II, V, VII, X, XI, XII or XIII), a membrane protein, an exon, an intracellular protein (*e.g.*, a cytoplasmic protein, a nuclear protein, an organellar protein such as a mitochondrial protein or lysosomal protein), an extracellular protein, a structural protein, a signaling

protein, a regulatory protein, a transport protein, a sensory protein, a motor protein, a defense protein, or a storage protein, an anti-inflammatory signaling molecules into cells for treatment of immune diseases, including but not limited to arthritis, psoriasis, lupus, coeliac disease, glomerulonephritis, hepatitis, and inflammatory bowel disease.

[298] The size of the inserted gene or nucleic acid can vary from about 1 bp to about 50,000 bp. In some embodiments, the size of the inserted gene or nucleic acid can be about 1 bp, 10 bp, 50 bp, 100 bp, 150 bp, 200 bp, 250 bp, 300 bp, 350 bp, 400 bp, 600 bp, 800 bp, 1000 bp, 1200 bp, 1400 bp, 1600 bp, 1800 bp, 2000 bp, 2200 bp, 2400 bp, 2600 bp, 2800 bp, 3000 bp, 3200 bp, 3400 bp, 3600 bp, 3800 bp, 4000 bp, 4200 bp, 4400 bp, 4600 bp, 4800 bp, 5000 bp, 5200 bp, 5400 bp, 5600 bp, 5800 bp, 6000 bp, 6200, 6400 bp, 6600 bp, 6800 bp, 7000 bp, 7200 bp, 7400 bp, 7600 bp, 7800 bp, 8000 bp, 8200 bp, 8400 bp, 8600 bp, 8800 bp, 9000 bp, 9200 bp, 9400 bp, 9600 bp, 9800 bp, 10,000 bp, 10,200 bp, 10,400 bp, 10,600 bp, 10,800 bp, 11,000 bp, 11,200 bp, 11,400 bp, 11,600 bp, 11,800 bp, 12,000 bp, 14,000 bp, 16,000 bp, 18,000 bp, 20,000 bp, 30,000 bp, 40,000 bp, 50,000 bp, or any range that is formed from any two of those values as endpoints.

[299] In some embodiments, the site-specific engineering using the gene of interest or nucleic acid sequence of interest disclosed herein is for the engineering of T cells and NKs for tumor targeting or allogeneic generation. These can involve the use of receptor or CAR for tumor specificity, anti-PD1 antibody, cytokines like IFN-gamma, TNF-alpha, IL-15, IL-12, IL-18, IL-21, and IL-10, and immune escape genes.

[300] In the present disclosure, the site-specific insertion of the gene of interest or nucleic acid of interest is performed through Programmable Addition via Site-Specific Targeting Elements (PASTE). Components for inserting a gene of interest or a nucleic acid of interest using PASTE are for example, without limitation, a nuclease, a gRNA adding the integration site, a DNA or RNA strand comprising the gene or nucleic acid linked to a sequence that is complementary or associated to the integration site, and an integration enzyme. Components for inserting a gene of interest or a nucleic acid of interest using PASTE are for example, without limitation, a prime editor expression, pegRNA adding the integration site, nicking guide RNA, integration enzyme (Cre or serine recombinase), transgene vector comprising the gene of interest or nucleic acid sequence of interest with gene and integration signal. The nuclease and prime editor integrate the integration site into the genome. The integration enzyme integrates the gene of interest into the integration site. In some embodiments, the transgene vector comprising the gene or nucleic acid sequence of interest with gene and

integration signal is a DNA minicircle devoid of bacterial DNA sequences. In some embodiments, the transgenic vector is a eukaryotic or prokaryotic vector.

[301] As used herein, the term "vector" or "transgene vector" refers to a recombinant DNA molecule containing a desired coding sequence and appropriate nucleic acid sequences necessary for the expression of the operably linked coding sequence in a host organism. Nucleic acid sequences necessary for expression in prokaryotes usually include for example, without limitation, a promoter, an operator (optional), a ribosome binding site, and/or other sequences. Eukaryotic cells are generally known to utilize promoters (constitutive, inducible or tissue specific), enhancers, and termination and polyadenylation signals, although some elements may be deleted and other elements added without sacrificing the necessary expression. The transgenic vector may encode the PE and the integration enzyme, linked to each other via a linker. The linker can be a cleavable linker. For example, transgenic vector encoding the PE and the integration enzyme, linked to each other via a linker is pCMV PE2 P2A Cre comprises SEQ ID NO: 73. In some embodiments, the linker can be a non-cleavable linker. In some embodiments the nuclease, prime editor, and/or integration enzyme can be encoded in different vectors.

[302] A method of inserting multiple genes or nucleic acid sequences of interest into a single site according to embodiments of the present disclosure is illustrated in FIG. 12. In some embodiments, multiplexing involves inserting multiple genes of interest in multiple loci using unique pegRNA as illustrated in FIG. 13 (Merrick, C. A. *et al.*, *ACS Synth. Biol.* 2018, 7, 299−310). The insertion of multiple genes of interest or nucleic acids of interest into a cell genome, referred herein as "multiplexing," is facilitated by incorporation of the complementary 5' integration site to the 5' end of the DNA or RNA comprising the first nucleic acid and 3' integration site to the 3' end of the DNA or RNA comprising the last nucleic acid. In some embodiments, the number of genome of interest or amino acid sequences of interest that are inserted into a cell genome using multiplexing can be about 1, 2, 3, 4, 5, 6, 7, 8, 9 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, or any range that is formed from any two of those values as endpoints.

[303] In some embodiments, multiplexing allows integration of for example, signaling cascade, over-expression of a protein of interest with its cofactor, insertion of multiple genes mutated in a neoplastic condition, or insertion of multiple CARs for treatment of cancer.

[304] In some embodiments, the integration sites may be inserted into the genome using non-prime editing methods such as rAAV mediated nucleic acid integration, TALENS and

ZFNs. A number of unique properties make AAV a promising vector for human gene therapy (Muzyczka, CURRENT TOPICS IN MICROBIOLOGY AND IMMUNOLOGY, 158:97-129 (1992)). Unlike other viral vectors, AAVs have not been shown to be associated with any known human disease and are generally not considered pathogenic. Wild type AAV is capable of integrating into host chromosomes in a site-specific manner M. Kotin *et al.*, PROC. NATL. ACAD. SCI, USA, 87:2211-2215 (1990); R.J. Samulski, EMBO 10(12):3941-3950 (1991)). Instead of creating a double-stranded DNA break, AAV stimulates endogenous homologous recombination to achieve the DNA modification. Further, transcription activator-like effector nucleases (TALENs) and Zinc-finger nucleases (ZFNs) for genome editing and introducing targeted DSBs. The specificity of TALENs arises from two polymorphic amino acids, the so-called repeat variable diresidues (RVDs) located at positions 12 and 13 of a repeated unit. TALENS are linked to FokI nucleases, which cleaves the DNA at the desired locations. ZFNs are artificial restriction enzymes for custom site-specific genome editing. Zinc fingers themselves are transcription factors, where each finger recognizes 3-4 bases. By mixing and matching these finger modules, researchers can customize which sequence to target.

[305] As used herein, the terms "administration," "introducing," or "delivery" into a cell, a tissue, or an organ of a plasmid, nucleic acids, or proteins for modification of the host genome refers to the transport for such administration, introduction, or delivery that can occur *in vivo*, *in vitro,* or *ex vivo*. Plasmids, DNA, or RNA for genetic modification can be introduced into cells by transfection, which is typically accomplished by chemical means (*e.g.*, calcium phosphate transfection, polyethyleneimine (PEI) Or lipofection), physical means (electroporation or microinjection), infection (this typically means the introduction of an infectious agent such as a virus (*e.g.*, a baculovirus expressing the AAV Rep gene)), transduction (in microbiology, this refers to the stable infection of cells by viruses, or the transfer of genetic material from one microorganism to another by viral factors (*e.g.*, bacteriophages)). Vectors for the expression of a recombinant polypeptide, protein or oligonucleotide may be obtained by physical means (*e.g.*, calcium phosphate transfection, electroporation, microinjection, or lipofection) in a cell, a tissue, an organ or a subject. The vector can be delivered by preparing the vector in a pharmaceutically acceptable carrier for the *in vitro*, *ex vivo,* or *in vivo* delivery to the carrier.

[306] As used herein, the term "transfection" refers to the uptake of an exogenous nucleic acid molecule by a cell. A cell is "transfected" when an exogenous nucleic acid has been

introduced into the cell membrane. The transfection can be a single transfection, co-transfection, or multiple transfection. Numerous transfection techniques are generally known in the art. See, for example, Graham *et al.* (1973) Virology, 52: 456. Such techniques can be used to introduce one or more exogenous nucleic acid molecules into a suitable host cell.

[307] In some embodiments, the exogenous nucleic acid molecule and/or other components for gene editing are combined and delivered in a single transfection. In other embodiments, the exogenous nucleic acid molecule and/or other components for gene editing are not combined and delivered in a single transfection. In some embodiments, exogenous nucleic acid molecule and/or other components for gene editing are combined and delivered in a single transfection to comprise for example, without limitation, a prime editing vector, a landing site such as a landing site containing pegRNA, a nicking guide such as a nicking guide for stimulating prime editing, an expression vector such as an expression vector for a corresponding integrase or recombinase, a minicircle DNA cargo such as a minicircle DNA cargo encoding for green fluorescent protein (GFP), any derivatives thereof, and any combinations thereof. In some embodiments, the gene of interest or amino acid sequence of interest can be introduced using liposomes. In some embodiments, the gene of interest or amino acid sequence of interest can be delivered using suitable vectors for instance, without limitation, plasmids and viral vectors. Examples of viral vectors include, without limitation, adeno-associated viruses (AAV), lentiviruses, adenoviruses, other viral vectors, derivatives thereof, or combinations thereof. The proteins and one or more guide RNAs can be packaged into one or more vectors, *e.g.*, plasmids or viral vectors. In some embodiments, the delivery is via nanoparticles or exosomes. For example, exosomes can be particularly useful in delivery RNA.

[308] In some embodiments, the prime editing inserts the landing site with efficiencies of at least about 1%, at least about 5%, at least about 10%, at least about 15%, at least about, at least about 20%, at least about 25%, at least about 30%, at least about 35%, at least about 40%, at least about 45%, or at least about 50%. In some embodiments, the prime editing inserts the landing site(s) with efficiencies of about 1%, about 2%, about 3%, about 4%, about 5%, about 6%, about 7%, about 8%, about 9%, about 10%, about 11%, about 12%, about 13%, about 14%, about 15%, about 16%, about 17%, about 18%, about 19%, about 20%, about 21%, about 22%, about 23%, about 24%, about 25%, about 26%, about 27%, about 28%, about 29%, about 30%, about 31%, about 32%, about 33%, about 34%, about 35%, about 36%, about 37%, about 38%, about 39%, about 40%, about 41%, about 42%,

about 43%, about 44%, about 45%, about 46%, about 47%, about 48%, about 49%, about 50%, or any range that is formed from any two of those values as endpoints.

Sequences

[309] Sequences of enzymes, guides, integration sites, and plasmids can be found in Table 4 below.

Table 4

SEQ ID NO / DESCRIPTION / SOURCE	**SEQUENCE**
SEQ ID NO: 1 Lox71 (Artificial sequence)	ATAACTTCGTATAATGTATGCTATACGAACGGTA
SEQ ID NO: 2 Lox66 (Artificial sequence)	TACCGTTCGTATAATGTATGCTATACGAAGTTAT
SEQ ID NO: 3 attB (Artificial sequence)	GGCCGGCTTGTCGACGACGGCGGTCTCCGTCGTCAGGATCATCCGG
SEQ ID NO: 4 attP (Artificial Sequence)	CCGGATGATCCTGACGACGGAGACCGCCGTCGTCGACAAGCCGGCC
SEQ ID NO: 5 attB-TT (Artificial Sequence)	GGCTTGTCGACGACGGCGTTCTCCGTCGTCAGGATCAT
SEQ ID NO: 6 attP-TT (Artificial Sequence)	GTGGTTTGTCTGGTCAACCACCGCGTTCTCAGTGGTGTACGGTACAAACCCA
SEQ ID NO: 7	GGCTTGTCGACGACGGCGAACTCCGTCGTCAGGATCAT

attB-AA (Artificial Sequence)	
SEQ ID NO: 8 attP-AA (Artificial Sequence)	GTGGTTTGTCTGGTCAACCACCGCGAACTCAGTGGTGTACGGTAC AAACCCA
SEQ ID NO: 9 attB-CC (Artificial Sequence)	GGCTTGTCGACGACGGCGCCCTCCGTCGTCAGGATCAT
SEQ ID NO: 10 attP-CC (Artificial Sequence)	GTGGTTTGTCTGGTCAACCACCGCGCCCTCAGTGGTGTACGGTACA AACCCA
SEQ ID NO: 11 attB-GG (Artificial Sequence)	GGCTTGTCGACGACGGCGGGCTCCGTCGTCAGGATCAT
SEQ ID NO: 12 attP-GG (Artificial Sequence)	GTGGTTTGTCTGGTCAACCACCGCGGGCTCAGTGGTGTACGGTAC AAACCCA
SEQ ID NO: 13 attB-TG (Artificial Sequence)	GGCTTGTCGACGACGGCGTGCTCCGTCGTCAGGATCAT
SEQ ID NO: 14 attP-TG (Artificial Sequence)	GTGGTTTGTCTGGTCAACCACCGCGTGCTCAGTGGTGTACGGTACA AACCCA
SEQ ID NO: 15 attB-GT (Artificial Sequence)	GGCTTGTCGACGACGGCGGTCTCCGTCGTCAGGATCAT

SEQ ID NO: 16 attP-GT (Artificial Sequence)	GTGGTTTGTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACA AACCCA
SEQ ID NO: 17 attB-CT (Artificial Sequence)	GGCTTGTCGACGACGGCGCTCTCCGTCGTCAGGATCAT
SEQ ID NO: 18 attP-CT (Artificial Sequence)	GTGGTTTGTCTGGTCAACCACCGCGCTCTCAGTGGTGTACGGTACA AACCCA
SEQ ID NO: 19 attB-CA (Artificial Sequence)	GGCTTGTCGACGACGGCGCACTCCGTCGTCAGGATCAT
SEQ ID NO: 20 attP-CA (Artificial Sequence)	GTGGTTTGTCTGGTCAACCACCGCGCACTCAGTGGTGTACGGTACA AACCCA
SEQ ID NO: 21 attB-TC (Artificial Sequence)	GGCTTGTCGACGACGGCGTCCTCCGTCGTCAGGATCAT
SEQ ID NO: 22 attP-TC (Artificial Sequence)	GTGGTTTGTCTGGTCAACCACCGCGTCCTCAGTGGTGTACGGTACA AACCCA
SEQ ID NO: 23 attB-GA (Artificial Sequence)	GGCTTGTCGACGACGGCGGACTCCGTCGTCAGGATCAT
SEQ ID NO: 24 attP-GA (Artificial Sequence)	GTGGTTTGTCTGGTCAACCACCGCGGACTCAGTGGTGTACGGTAC AAACCCA

SEQ ID NO: 25 attB-AG (Artificial Sequence)	GGCTTGTCGACGACGGCGAGCTCCGTCGTCAGGATCAT
SEQ ID NO: 26 attP-AG (Artificial Sequence)	GTGGTTTGTCTGGTCAACCACCGCGAGCTCAGTGGTGTACGGTAC AAACCCA
SEQ ID NO: 27 attB-AC (Artificial Sequence)	GGCTTGTCGACGACGGCGACCTCCGTCGTCAGGATCAT
SEQ ID NO: 28 attP-AC (Artificial Sequence)	GTGGTTTGTCTGGTCAACCACCGCGACCTCAGTGGTGTACGGTACA AACCCA
SEQ ID NO: 29 attB-AT (Artificial Sequence)	GGCTTGTCGACGACGGCGATCTCCGTCGTCAGGATCAT
SEQ ID NO: 30 attP-AT (Artificial Sequence)	GTGGTTTGTCTGGTCAACCACCGCGATCTCAGTGGTGTACGGTACA AACCCA
SEQ ID NO: 31 attB-GC (Artificial Sequence	GGCTTGTCGACGACGGCGGCCTCCGTCGTCAGGATCAT
SEQ ID NO: 32 attP-GC (Artificial Sequence)	GTGGTTTGTCTGGTCAACCACCGCGGCCTCAGTGGTGTACGGTACA AACCCA
SEQ ID NO: 33 attB-CG (Artificial Sequence)	GGCTTGTCGACGACGGCGCGCTCCGTCGTCAGGATCAT

SEQ ID NO: 34 attP-CG (Artificial Sequence)	GTGGTTTGTCTGGTCAACCACCGCGCGCTCAGTGGTGTACGGTACAAACCCA
SEQ ID NO: 35 attB-TA (Artificial Sequence)	GGCTTGTCGACGACGGCGTACTCCGTCGTCAGGATCAT
SEQ ID NO: 36 attP-TA (Artificial Sequence)	GTGGTTTGTCTGGTCAACCACCGCGTACTCAGTGGTGTACGGTACAAACCCA
SEQ ID NO: 37 C31-attB (Artificial Sequence)	TGCGGGTGCCAGGGCGTGCCCTTGGGCTCCCCGGGCGCGTACTCC
SEQ ID NO: 38 C31-attP (Artificial Sequence)	GTGCCCCAACTGGGGTAACCTTTGAGTTCTCTCAGTTGGGGG
SEQ ID NO: 39 R4-attB (Artificial Sequence)	GCGCCCAAGTTGCCCATGACCATGCCGAAGCAGTGGTAGAAGGGCACCGGCAGACAC
SEQ ID NO: 40 R4-attP (Artificial Sequence)	AGGCATGTTCCCCAAAGCGATACCACTTGAAGCAGTGGTACTGCTTGTGGGTACACTCTGCGGGTGATGA
SEQ ID NO: 41 BT1-attB (Artificial Sequence)	GTCCTTGACCAGGTTTTTGACGAAAGTGATCCAGATGATCCAGCTCCACACCCCGAACGC
SEQ ID NO: 42 BT1-attP (Artificial Sequence)	GGTGCTGGGTTGTTGTCTCTGGACAGTGATCCATGGGAAACTACTCAGCACCACCAATGTTCC

SEQ ID NO: 43 Bxb-attB (Artificial Sequence)	TCGGCCGGCTTGTCGACGACGGCGGTCTCCGTCGTCAGGATCATCCGGGC
SEQ ID NO: 44 Bxb-attP (Artificial Sequence)	GTCGTGGTTTGTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCCCGAC
SEQ ID NO: 45 TG1-attB (Artificial Sequence)	GATCAGCTCCGCGGGCAAGACCTTCTCCTTCACGGGGTGGAAGGTC
SEQ ID NO: 46 TG1-attP (Artificial Sequence)	TCAACCCCGTTCCAGCCCAACAGTGTTAGTCTTTGCTCTTACCCAGTTGGGCGGGATAGCCTGCCCG
SEQ ID NO: 47 C1-attB (Artificial Sequence)	AACGATTTTCAAAGGATCACTGAATCAAAAGTATTGCTCATCCACGCGAAATTTTTC
SEQ ID NO: 48 C1-attP (Artificial Sequence)	AATATTTTAGGTATATGATTTTGTTTATTAGTGTAAATAACACTATGTACCTAAAAT
SEQ ID NO: 49 C370-attB (Artificial Sequence)	TGTAAAGGAGACTGATAATGGCATGTACAACTATACTCGTCGGTAAAAAGGCA
SEQ ID NO: 50 C370-attP (Artificial Sequence)	TAAAAAAATACAGCGTTTTTCATGTACAACTATACTAGTTGTAGTGCCTAAA
SEQ ID NO: 51 K38-attB (Artificial Sequence)	GAGCGCCGGATCAGGGAGTGGACGGCCTGGGAGCGCTACACGCTGTGGCTGCGGTC

SEQ ID NO: 52 K38-attP (Artificial Sequence)	CCCTAATACGCAAGTCGATAACTCTCCTGGGAGCGTTGACAACTT GCGCACCCTGA
SEQ ID NO: 53 RB-attB (Artificial Sequence)	TCTCGTGGTGGTGGAAGGTGTTGGTGCGGGGTTGGCCGTGGTCGA GGTGGGGTGGTGGTAGCCATTCG
SEQ ID NO: 54 RV-attP (Artificial Sequence)	GCACAGGTGTAGTGTATCTCACAGGTCCACGGTTGGCCGTGGACT GCTGAAGAACATTCCACGCCAGGA
SEQ ID NO: 55 SPBC-attB (Artificial Sequence)	AGTGCAGCATGTCATTAATATCAGTACAGATAAAGCTGTATCTCCT GTGAACACAATGGGTGCCA
SEQ ID NO: 56 SPBC-attP (Artificial Sequence)	AAAGTAGTAAGTATCTTAAAAAACAGATAAAGCTGTATATTAAGA TACTTACTAC
SEQ ID NO: 57 TP901-attB (Artificial Sequence)	TGATAATTGCCAACACAATTAACATCTCAATCAAGGTAAATGCTTT TTCGTTTT
SEQ ID NO: 58 TP901-attP (Artificial Sequence)	AATTGCGAGTTTTTATTTCGTTTATTTCAATTAAGGTAACTAAAAA ACTCCTTT
SEQ ID NO: 59 Wβ-attB (Artificial Sequence)	AAGGTAGCGTCAACGATAGGTGTAACTGTCGTGTTTGTAACGGTA CTTCCAACAGCTGGCGTTTCAGT
SEQ ID NO: 60 Wβ-attP (Artificial Sequence)	TAGTTTTAAAGTTGGTTATTAGTTACTGTGATATTTATCACGGTAC CCAATAACCAATGAATATTTGA

SEQ ID NO: 61 A118-attB (Artificial Sequence)	TGTAACTTTTTCGGATCAAGCTATGAAGGACGCAAAGAGGGAACT AAACACTTAATT
SEQ ID NO: 62 A118-attP (Artificial Sequence)	TTGTTTAGTTCCTCGTTTTCTCTCGTTGGAAGAAGAAGAAACGAGA AACTAAAATTA
SEQ ID NO: 63 BL3-attB (Artificial Sequence)	CAACCTGTTGACATGTTTCCACAGACAACTCACGTGGAGGTAGTC ACGGCTTTTACGTTAGTT
SEQ ID NO: 64 BL3-attP (Artificial Sequence)	GAGAATACTGTTGAACAATGAAAAACTAGGCATGTAGAAGTTGTT TGTGCACTAACTTTAA
SEQ ID NO: 65 MR11-attB (Artificial Sequence)	ACAGGTCAACACATCGCAGTTATCGAACAATCTTCGAAAATGTAT GGAGGCACTTGTATCAATATAGGATGTATACCTTCGAAGACACTT GTACATGATGGATTAGAAGGCAAATCCTTT
SEQ ID NO: 66 MR11-attP (Artificial Sequence)	CAAAATAAAAAACATTGATTTTTATTAACTTCTTTTGTGCGGAACT ACGAACAGTTCATTAATACGAAGTGTACAAACTTCCATACAAAAA TAACCACGACAATTAAGACGTGGTTTCTA
SEQ ID NO: 67 attL (Artificial Sequence)	ATTATTTCTCACCCTGA
SEQ ID NO: 68 attR (Artificial Sequence)	ATCATCTCCCACCCGGA
SEQ ID NO: 69 Vox (Artificial Sequence)	AATAGGTCTG AGAACGCCCA TTCTCAGACG TATT

SEQ ID NO: 70 FRT (Artificial Sequence)	GAAGTTCCTATAC TTTCTAGA GAATAGGAACTTC
SEQ ID NO: 71 Cre recombinase expression plasmid (Artificial Sequence)	GGTCGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGG GGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACT TACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCC ATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGG GACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCC CACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTA TTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGT ACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATT AGTCATCGCTATTACCATGGTCGAGGTGAGCCCCACGTTCTGCTTC ACTCTCCCCATCTCCCCCCCCTCCCCACCCCCAATTTTGTATTTATT TATTTTTTAATTATTTTGTGCAGCGATGGGGGCGGGGGGGGGGGG GGCGCGCGCCAGGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG GGGGGGGCGGGGGGGGGGCGGCGGCAGCCAATCAGAGCGGCGCGC TCCGAAAGTTTCCTTTTATGGCGAGGCGGCGGCGGCGGCGGCCCT ATAAAAAGCGAAGCGCGCGGCGGGCGGGAGTCGCTGCGCGCTGC CTTCGCCCCGTGCCCCGCTCCGCCGCCGCCTCGCGCCGCCCGCCCC GGCTCTGACTGACCGCGTTACTCCCACAGGTGAGCGGGCGGGACG GCCCTTCTCCTCCGGGCTGTAATTAGCGCTTGGTTTAATGACGGCT TGTTTCTTTTCTGTGGCTGCGTGAAAGCCTTGAGGGGCTCCGGGAG GGCCCTTTGTGCGGGGGGAGCGGCTCGGGGGGTGCGTGCGTGTGT GTGTGCGTGGGGAGCGCCGCGTGCGGCTCCGCGCTGCCCGGCGGC TGTGAGCGCTGCGGGCGCGGCGCGGGGCTTTGTGCGCTCCGCAGT GTGCGCGAGGGGAGCGCGGCCGGGGGCGGTGCCCCGCGGTGCGG GGGGGGCTGCGAGGGGAACAAAGGCTGCGTGCGGGGTGTGTGCG TGGGGGGGTGAGCAGGGGGTGTGGGCGCGTCGGTCGGGCTGCAA CCCCCCCTGCACCCCCCTCCCCGAGTTGCTGAGCACGGCCCGGCTT CGGGTGCGGGGCTCCGTACGGGGCGTGGCGCGGGGCTCGCCGTGC CGGGCGGGGGGTGGCGGCAGGTGGGGGTGCCGGGCGGGGCGGGG CCGCCTCGGGCCGGGGAGGGCTCGGGGGAGGGGCGCGGCGGCCC CCGGAGCGCCGGCGGCTGTCGAGGCGCGGCGAGCCGCAGCCATTG CCTTTTATGGTAATCGTGCGAGAGGGCGCAGGGACTTCCTTTGTCC CAAATCTGTGCGGAGCCGAAATCTGGGAGGCGCCGCCGCACCCCC TCTAGCGGGCGCGGGGCGAAGCGGTGCGGCGCCGGCAGGAAGGA AATGGGCGGGGAGGGCCTTCGTGCGTCGCCGCGCCGCCGTCCCCT TCTCCCTCTCCAGCCTCGGGGCTGTCCGCGGGGGGACGGCTGCCTT CGGGGGGGACGGGGCAGGGCGGGGTTCGGCTTCTGGCGTGTGACC GGCGGCTCTAGAGCCTCTGCTAACCATGTTCATGCCTTCTTCTTTTT CCTACAGCTCCTGGGCAACGTGCTGGTTATTGTGCTGTCTCATCAT TTTGGCAAAGAATTCTGAGCCGCCACCATGGCCAATTTACTGACC GTACACCAAAATTTGCCTGCATTACCGGTCGATGCAACGAGTGAT GAGGTTCGCAAGAACCTGATGGACATGTTCAGGGATCGCCAGGCG TTTTCTGAGCATACCTGGAAAATGCTTCTGTCCGTTTGCCGGTCGT GGGCGGCATGGTGCAAGTTGAATAACCGGAAATGGTTTCCCGCAG AACCTGAAGATGTTCGCGATTATCTTCTATATCTTCAGGCGCGCGG TCTGGCAGTAAAAACTATCCAGCAACATTTGGGCCAGCTAAACAT GCTTCATCGTCGGTCCGGGCTGCCACGACCAAGTGACAGCAATGC TGTTTCACTGGTTATGCGGCGGATCCGAAAAGAAAACGTTGATGC

	CGGTGAACGTGCAAAACAGGCTCTAGCGTTCGAACGCACTGATTT CGACCAGGTTCGTTCACTCATGGAAAATAGCGATCGCTGCCAGGA TATACGTAATCTGGCATTTCTGGGGATTGCTTATAACACCCTGTTA CGTATAGCCGAAATTGCCAGGATCAGGGTTAAAGATATCTCACGT ACTGACGGTGGGAGAATGTTAATCCATATTGGCAGAACGAAAACG CTGGTTAGCACCGCAGGTGTAGAGAAGGCACTTAGCCTGGGGGTA ACTAAACTGGTCGAGCGATGGATTTCCGTCTCTGGTGTAGCTGATG ATCCGAATAACTACCTGTTTTGCCGGGTCAGAAAAAATGGTGTTG CCGCGCCATCTGCCACCAGCCAGCTATCAACTCGCGCCCTGGAAG GGATTTTTGAAGCAACTCATCGATTGATTTACGGCGCTAAGGATG ACTCTGGTCAGAGATACCTGGCCTGGTCTGGACACAGTGCCCGTG TCGGAGCCGCGCGAGATATGGCCCGCGCTGGAGTTTCAATACCGG AGATCATGCAAGCTGGTGGCTGGACCAATGTAAATATTGTCATGA ACTATATCCGTAACCTGGATAGTGAAACAGGGGCAATGGTGCGCC TGCTGGAAGATGGCGATGGACCGGTGGAACAAAAACTTATTTCTG AAGAAGATCTGTGATAGCGGCCGCACTCCTCAGGTGCAGGCTGCC TATCAGAAGGTGGTGGCTGGTGTGGCCAATGCCCTGGCTCACAAA TACCACTGAGATCTTTTTCCCTCTGCCAAAAATTATGGGGACATCA TGAAGCCCCTTGAGCATCTGACTTCTGGCTAATAAAGGAAATTTAT TTTCATTGCAATAGTGTGTTGGAATTTTTTGTGTCTCTCACTCGGAA GGACATATGGGAGGGCAAATCATTTAAAACATCAGAATGAGTATT TGGTTTAGAGTTTGGCAACATATGCCCATATGCTGGCTGCCATGAA CAAAGGTTGGCTATAAAGAGGTCATCAGTATATGAAACAGCCCCC TGCTGTCCATTCCTTATTCCATAGAAAAGCCTTGACTTGAGGTTAG ATTTTTTTTATATTTTGTTTTGTGTTATTTTTTTCTTTAACATCCCTA AAATTTTCCTTACATGTTTTACTAGCCAGATTTTTCCTCCTCTCCTG ACTACTCCCAGTCATAGCTGTCCCTCTTCTCTTATGGAGATCCCTC GACCTGCAGCCCAAGCTTGGCGTAATCATGGTCATAGCTGTTTCCT GTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCC GGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAA CTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAA ACCTGTCGTGCCAGCGGATCCGCATCTCAATTAGTCAGCAACCAT AGTCCCGCCCCTAACTCCGCCCATCCCGCCCCTAACTCCGCCCAGT TCCGCCCATTCTCCGCCCCATGGCTGACTAATTTTTTTTATTTATGC AGAGGCCGAGGCCGCCTCGGCCTCTGAGCTATTCCAGAAGTAGTG AGGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAAAGCTAACTTGT TTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAA ATTTCACAAATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTTG TCCAAACTCATCAATGTATCTTATCATGTCTGGATCCGCTGCATTA ATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCG CTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGG CTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTA TCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAA AGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGG CGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCG ACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGAT ACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCC GACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGA AGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGG TGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGT TCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCC AACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGT AACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTC

	TTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTT GGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTT GGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGT TTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCT CAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGA ACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAA GGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATC AATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATG CTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCA TCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGG AGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACC CACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCG GAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCA TCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCC AGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTG GTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCC AACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAG CGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGC CGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTT ACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACT CAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCT CTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAA CTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAAC TCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCAC TCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTT CTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGA ATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTC AATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATA CATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCG CACATTTCCCCGAAAAGTGCCACCTG
SEQ ID NO: 72 GFP-Lox66 Cre expression plasmid (Artificial Sequence)	AGCTCTGATCAAGAGACAGGATGAGGATCGTTTCGCATGATTGAA CAAGATGGATTGCACGCAGGTTCTCCGGCCGCTTGGGTGGAGAGG CTATTCGGCTATGACTGGGCACAACAGACAATCGGCTGCTCTGAT GCCGCCGTGTTCCGGCTGTCAGCGCAGGGGCGCCCGGTTCTTTTTG TCAAGACCGACCTGTCCGGTGCCCTGAATGAACTGCAAGACGAGG CAGCGCGGCTATCGTGGCTGGCCACGACGGGCGTTCCTTGCGCAG CTGTGCTCGACGTTGTCACTGAAGCGGGAAGGGACTGGCTGCTAT TGGGCGAAGTGCCGGGGCAGGATCTCCATGTCATCTACACCTTGC TCCTGCCGAGAAAGTATCCATCATGGCTGATGCAATGCGGCGGCT GCATACGCTTGATCCGGCTACCTGCCCATTCGACCACCAAGCGAA ACATCGCATCGAGCGAGCACGTACTCGGATGGAAGCCGGTCTTGT CGATCAGGATGATCTGGACGAAGAGCATCAGGGGCTCGCGCCAGC CGAACTGTTCGCCAGGCTCAAGGCGAGCATGCCCGACGGCGAGGA TCTCGTCGTGACCCATGGCGATGCCTGCTTGCCGAATATCATGGTG GAAAATGGCCGCTTTTCTGGATTCATCGACTGTGGCCGGCTGGGTG TGGCGGACCGCTATCAGGACATAGCGTTGGCTACCCGTGATATTG CTGAAGAGCTTGGCGGCGAATGGGCTGACCGCTTCCTCGTGCTTTA CGGTATCGCCGCTCCCGATTCGCAGCGCATCGCCTTCTATCGCCTT CTTGACGAGTTCTTCTGAATTATTAACTCGAGATCCACTAGAGTGT GGCGGCCGCATTCTTATAATCAGCATCATGATGTGGTACCACATCA TGATGCTGATTACCCCCAACTGAGAGAACTCAAAGGTTACCCCAG TTGGGGCGGGCCCACAAATAAAGCAATAGCATCACAAATTTCACA

AATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAAC
TCATCGAGCTCGAGATCTGGCGAAGGCGATGGGGGTCTTGAAGGC
GTGCTGGTACTCCACGATGCCCAGCTCGGTGTTGCTGTGCAGCTCC
TCCACGCGGCGGAAGGCGAACATGGGGCCCCCGTTCTGCAGGATG
CTGGGGTGGATGGCGCTCTTGAAGTGCATGTGGCTGTCCACCACG
AAGCTGTAGTAGCCGCCGTCGCGCAGGCTGAAGGTGCGGGCGAAG
CTGCCCACCAGCACGTTATCGCCCATGGGGTGCAGGTGCTCCACG
GTGGCGTTGCTGCGGATGATCTTGTCGGTGAAGATCACGCTGTCCT
CGGGGAAGCCGGTGCCCACCACCTTGAAGTCGCCGATCACGCGGC
CGGCCTCGTAGCGGTAGCTGAAGCTCACGTGCAGCACGCCGCCGT
CCTCGTACTTCTCGATGCGGGTGTTGGTGTAGCCGCCGTTGTTGAT
GGCGTGCAGGAAGGGGTTCTCGTAGCCGCTGGGGTAGGTGCCGAA
GTGGTAGAAGCCGTAGCCCATCACGTGGCTCAGCAGGTAGGGGCT
GAAGGTCAGGGCGCCTTTGGTGCTCTTCATCTTGTTGGTCATGCGG
CCCTGCTCGGGGGTGCCCTCTCCGCCGCCCACCAGCTCGAACTCCA
CGCCGTTCAGGGTGCCGGTGATGCGGCACTCGATCTTCATGGCGG
GCATGGTGGCGACCGGTAGCGCTAGCGGCTTCGGATAACTTCGTA
TAGCATACATTATACGAACGGTAAGCGCTACCGCCGGCATACCCA
AGTGAAGTTGCTCGCAGCTTATAGTCGCGCCCGGGGAGCCCAAGG
GCACGCCCTGGCACCGCGGCCGCTGAGTCTCGACCATCATCATCA
TCATCATTGAGTTTATCTGGGATAACAGGGTAATGTCATCTAGGGA
TAACAGGGTATGTCATCTGGGATAACAGGGTAATGTATCTAGGGA
TAACAGGGTAATGTCATCTGGGATAACAGGGTAATGTCATCTAGG
GATAACAGGGTATGTCATCTGGGATAACAGGGTAATGTATCTAGG
GATAACAGGGTAATGTCATCTGGGATAACAGGGTAATGTCATCTA
GGGATAACAGGGTATGTCATCTGGGATAACAGGGTAATGTATCTA
GGGATAACAGGGTAATGTCATCTGGGATAACAGGGTAATGTCATC
TAGGGATAACAGGGTATGTCATCTGGGATAACAGGGTAATGTATC
TAGGGATAACAGGGTAATGTCATCTGGGATAACAGGGTAATGTCA
TCTAGGGATAACAGGGTATGTCATCTGGGATAACAGGGTAATGTA
TCTAGGGATAACAGGGTAATGTCATCTGGGATAACAGGGTAATGT
CATCTAGGGATAACAGGGTATGTCATCTGGGATAACAGGGTAATG
TATCTAGGGATAACAGGGTAATGTCATCTGGGATAACAGGGTAAT
GTCATCTAGGGATAACAGGGTAAATGTCATCTAGGGATAACAGGG
TAATGTCATCTAGGGATAACAGGGTAATGTCATCTGGGATAACAG
GGTAATGTCATCTAGGGATAACAGGGTAATGTATCGCCAGCGTCG
CACAGCATGTTTGCTTGTCGCCGTCGCGTCTGTCACATCTTTTCCG
CCAGCAGTTAGGGATTAGCGTCTTAAGCTGGCGCGAGGACCAACG
TATCAGCCAGGCGAAGCTGCTTTTGAGCACCACCCGGATGCCTAT
CGCCACCGTCGGTCGCAATGTTGGTTTTGACGATCAACTCTATTTC
TCGCGGGTATTTAAAAAATGCACCGGGGCCAGCCCGAGCGAGTTC
CGTGCCGGTTGTGAAGAAAAAGTGAATGATGTAGCCGTCAAGTTG
TCATAATTGGTAACGAATCAGACAATTGACGGCTTGACGGAGTAG
CATAGGGTTTGCAGAATCCCTGCTTCGTCCATTTGACAGGCACATT
ATGCATGCCGCTTCGCCTTCGCGCGCGAATTGATCTGCTGCCTCGC
GCGTTTCGGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCC
GGAGACGGTCACAGCTTGTCTGTAAGCGGATGCCGGGAGCAGACA
AGCCCGTCAGGGCGCGTCAGCGGGTGTTGGCGGGTGTCGGGGCGC
AGCCATGACCCAGTCACGTAGCGATAGCGGAGTGTATACTGGCTT
AACTATGCGGCATCAGAGCAGATTGTACTGAGAGTGCACCATATG
CGGTGTGAAATACCGCACAGATGCGTAAGGAGAAAATACCGCATC
AGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCG
TTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATAC

	GGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGA GCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTT GCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAA AATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAA AGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTG TTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCG GGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTT CGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCC CCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGA GTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCAC TGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGA GTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGT ATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGA GTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGT GGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAGGA TCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGT GGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGCGGATACA TATTTGAATGTATTTAGAAAAATAAACAAAAGAGTTTGTAGAAAC GCAAAAAGGCCATCCGTCAGGATGGCCTTCTGCTTAATTTGATGCC TGGCAGTTTATGGCGGGCGTCCTGCCCGCCACCCTCCGGGCCGTTG CTTCGCAACGTTCAAATCCGCTCCCGGCGGATTTGTCCTACTCAGG AGAGCGTTCACCGACAAACAACAGATAAAACGAAAGGCCCAGTC TTTCGACTGAGCCTTTCGTTTTATTTGATGCCTGGCAGTTCCCTACT CTCGCATGGGGAGACCCCACACTACCATCGGCGCTACGGCGTTTC ACTTCTGAGTTCGGCATGGGGTCAGGTGGGACCACCGCGCTACTG CCGCCAGGCAAATTCTGTTTTATCAGACCGCTTCTGCGTTCTGATT TAATCTGTATCAGGCTGAAAATCTTCTCTCATCCGCCAAAACAGCC AAGCTGGAGACCGTTTGGCCCCCCTCGAGCACGTAGAAAGCCAGT CCGCAGAAACGGTGCTGACCCCGGATGAATGTCAGCTACTGGGCT ATCTGGACAAGGGAAAACGCAAGCGCAAAGAGAAAGCAGGTAGC TTGCAGTGGGCTTACATGGCGATAGCTAGACTGGGCGGTTTTATG GACAGCAAGCGAACCGGAATTGCCAGCTGGGGCGCCCTCTGGTAA GGTTGGGAAGCCCTGCAAAGTAAACTGGATGGCTTTCTCGCCGCC AAGGATCTGATGGCGCAGGGGATCA
SEQ ID NO: 73 pCMV PE2 P2A Cre plasmid (Artificial Sequence)	ACGCGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTA CGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACATA ACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCG CCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATA GGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACT GCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCC CCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCC AGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGT ATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATC AATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTC CACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAAC GGGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAA TGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTG GTTTAGTGAACCGTCAGATCCGCTAGAGATCCGCGGCCGCTAATA CGACTCACTATAGGGAGAGCCGCCACCATGAAACGGACAGCCGAC GGAAGCGAGTTCGAGTCACCAAAGAAGAAGCGGAAAGTCGACAA GAAGTACAGCATCGGCCTGGACATCGGCACCAACTCTGTGGGCTG GGCCGTGATCACCGACGAGTACAAGGTGCCCAGCAAGAAATTCAA

	GGTGCTGGGCAACACCGACCGGCACAGCATCAAGAAGAACCTGAT CGGAGCCCTGCTGTTCGACAGCGGCGAAACAGCCGAGGCCACCCG GCTGAAGAGAACCGCCAGAAGAAGATACACCAGACGGAAGAACC GGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGATGGCCAAGG TGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTCCTGGTGG AAGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCG TGGACGAGGTGGCCTACCACGAGAAGTACCCCACCATCTACCACC TGAGAAAGAAACTGGTGGACAGCACCGACAAGGCCGACCTGCGG CTGATCTATCTGGCCCTGGCCCACATGATCAAGTTCCGGGGCCACT TCCTGATCGAGGGCGACCTGAACCCCGACAACAGCGACGTGGACA AGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGG AAAACCCCATCAACGCCAGCGGCGTGGACGCCAAGGCCATCCTGT CTGCCAGACTGAGCAAGAGCAGACGGCTGGAAAATCTGATCGCCC AGCTGCCCGGCGAGAAGAAGAATGGCCTGTTCGGAAACCTGATTG CCCTGAGCCTGGGCCTGACCCCCAACTTCAAGAGCAACTTCGACC TGGCCGAGGATGCCAAACTGCAGCTGAGCAAGGACACCTACGACG ACGACCTGGACAACCTGCTGGCCCAGATCGGCGACCAGTACGCCG ACCTGTTTCTGGCCGCCAAGAACCTGTCCGACGCCATCCTGCTGAG CGACATCCTGAGAGTGAACACCGAGATCACCAAGGCCCCCCTGAG CGCCTCTATGATCAAGAGATACGACGAGCACCACCAGGACCTGAC CCTGCTGAAAGCTCTCGTGCGGCAGCAGCTGCCTGAGAAGTACAA AGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGCCGGCTACAT TGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCC CATCCTGGAAAAGATGGACGGCACCGAGGAACTGCTCGTGAAGCT GAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTCGACAACG GCAGCATCCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTC TGCGGCGGCAGGAAGATTTTTACCCATTCCTGAAGGACAACCGGG AAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACTACGTGG GCCCTCTGGCCAGGGGAAACAGCAGATTCGCCTGGATGACCAGAA AGAGCGAGGAAACCATCACCCCCTGGAACTTCGAGGAAGTGGTGG ACAAGGGCGCTTCCGCCCAGAGCTTCATCGAGCGGATGACCAACT TCGATAAGAACCTGCCCAACGAGAAGGTGCTGCCCAAGCACAGCC TGCTGTACGAGTACTTCACCGTGTATAACGAGCTGACCAAAGTGA AATACGTGACCGAGGGAATGAGAAAGCCCGCCTTCCTGAGCGGCG AGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGA AAGTGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAATC GAGTGCTTCGACTCCGTGGAAATCTCCGGCGTGGAAGATCGGTTC AACGCCTCCCTGGGCACATACCACGATCTGCTGAAAATTATCAAG GACAAGGACTTCCTGGACAATGAGGAAAACGAGGACATTCTGGA AGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGATGAT CGAGGAACGGCTGAAAACCTATGCCCACCTGTTCGACGACAAAGT GATGAAGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGC TGAGCCGGAAGCTGATCAACGGCATCCGGGACAAGCAGTCCGGCA AGACAATCCTGGATTTCCTGAAGTCCGACGGCTTCGCCAACAGAA ACTTCATGCAGCTGATCCACGACGACAGCCTGACCTTTAAAGAGG ACATCCAGAAAGCCCAGGTGTCCGGCCAGGGCGATAGCCTGCACG AGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCA TCCTGCAGACAGTGAAGGTGGTGGACGAGCTCGTGAAAGTGATGG GCCGGCACAAGCCCGAGAACATCGTGATCGAAATGGCCAGAGAG AACCAGACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGAAT GAAGCGGATCGAAGAGGGCATCAAAGAGCTGGGCAGCCAGATCC TGAAAGAACACCCCGTGGAAAACACCCAGCTGCAGAACGAGAAG CTGTACCTGTACTACCTGCAGAATGGGCGGGATATGTACGTGGAC

	CAGGAACTGGACATCAACCGGCTGTCCGACTACGATGTGGACGCT ATCGTGCCTCAGAGCTTTCTGAAGGACGACTCCATCGACAACAAG GTGCTGACCAGAAGCGACAAGAACCGGGGCAAGAGCGACAACGT GCCCTCCGAAGAGGTCGTGAAGAAGATGAAGAACTACTGGCGGC AGCTGCTGAACGCCAAGCTGATTACCCAGAGAAAGTTCGACAATC TGACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTGGATAAGGCC GGCTTCATCAAGAGACAGCTGGTGGAAACCCGGCAGATCACAAAG CACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTACGAC GAGAATGACAAGCTGATCCGGGAAGTGAAAGTGATCACCCTGAA GTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGTTTTACAAA GTGCGCGAGATCAACAACTACCACCACGCCCACGACGCCTACCTG AACGCCGTCGTGGGAACCGCCCTGATCAAAAAGTACCCTAAGCTG GAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGACGTGCGG AAGATGATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGC CAAGTACTTCTTCTACAGCAACATCATGAACTTTTTCAAGACCGAG ATTACCCTGGCCAACGGCGAGATCCGGAAGCGGCCTCTGATCGAG ACAAACGGCGAAACCGGGGAGATCGTGTGGGATAAGGGCCGGGA TTTTGCCACCGTGCGGAAAGTGCTGAGCATGCCCCAAGTGAATAT CGTGAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGT CTATCCTGCCCAAGAGGAACAGCGATAAGCTGATCGCCAGAAAGA AGGACTGGGACCCTAAGAAGTACGGCGGCTTCGACAGCCCCACCG TGGCCTATTCTGTGCTGGTGGTGGCCAAAGTGGAAAAGGGCAAGT CCAAGAAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCA TGGAAAGAAGCAGCTTCGAGAAGAATCCCATCGACTTTCTGGAAG CCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTG CCTAAGTACTCCCTGTTCGAGCTGGAAAACGGCCGGAAGAGAATG CTGGCCTCTGCCGGCGAACTGCAGAAGGGAAACGAACTGGCCCTG CCCTCCAAATATGTGAACTTCCTGTACCTGGCCAGCCACTATGAGA AGCTGAAGGGCTCCCCCGAGGATAATGAGCAGAAACAGCTGTTTG TGGAACAGCACAAGCACTACCTGGACGAGATCATCGAGCAGATCA GCGAGTTCTCCAAGAGAGTGATCCTGGCCGACGCTAATCTGGACA AAGTGCTGTCCGCCTACAACAAGCACCGGGATAAGCCCATCAGAG AGCAGGCCGAGAATATCATCCACCTGTTTACCCTGACCAATCTGG GAGCCCCTGCCGCCTTCAAGTACTTTGACACCACCATCGACCGGA AGAGGTACACCAGCACCAAAGAGGTGCTGGACGCCACCCTGATCC ACCAGAGCATCACCGGCCTGTACGAGACACGGATCGACCTGTCTC AGCTGGGAGGTGACTCTGGAGGATCTAGCGGAGGATCCTCTGGCA GCGAGACACCAGGAACAAGCGAGTCAGCAACACCAGAGAGCAGT GGCGGCAGCAGCGGCGGCAGCAGCACCCTAAATATAGAAGATGA GTATCGGCTACATGAGACCTCAAAAGAGCCAGATGTTTCTCTAGG GTCCACATGGCTGTCTGATTTTCCTCAGGCCTGGGCGGAAACCGG GGGCATGGGACTGGCAGTTCGCCAAGCTCCTCTGATCATACCTCTG AAAGCAACCTCTACCCCCGTGTCCATAAAACAATACCCCATGTCA CAAGAAGCCAGACTGGGGATCAAGCCCCACATACAGAGACTGTTG GACCAGGGAATACTGGTACCCTGCCAGTCCCCCTGGAACACGCCC CTGCTACCCGTTAAGAAACCAGGGACTAATGATTATAGGCCTGTC CAGGATCTGAGAGAAGTCAACAAGCGGGTGGAAGACATCCACCC CACCGTGCCCAACCCTTACAACCTCTTGAGCGGGCTCCCACCGTCC CACCAGTGGTACACTGTGCTTGATTTAAAGGATGCCTTTTTCTGCC TGAGACTCCACCCCACCAGTCAGCCTCTCTTCGCCTTTGAGTGGAG AGATCCAGAGATGGGAATCTCAGGACAATTGACCTGGACCAGACT CCCACAGGGTTTCAAAAACAGTCCCACCCTGTTTAATGAGGCACT GCACAGAGACCTAGCAGACTTCCGGATCCAGCACCCAGACTTGAT

	CCTGCTACAGTACGTGGATGACTTACTGCTGGCCGCCACTTCTGAG CTAGACTGCCAACAAGGTACTCGGGCCCTGTTACAAACCCTAGGG AACCTCGGGTATCGGGCCTCGGCCAAGAAAGCCCAAATTTGCCAG AAACAGGTCAAGTATCTGGGGTATCTTCTAAAAGAGGGTCAGAGA TGGCTGACTGAGGCCAGAAAAGAGACTGTGATGGGGCAGCCTACT CCGAAGACCCCTCGACAACTAAGGGAGTTCCTAGGGAAGGCAGGC TTCTGTCGCCTCTTCATCCCTGGGTTTGCAGAAATGGCAGCCCCCC TGTACCCTCTCACCAAACCGGGGACTCTGTTTAATTGGGGCCCAGA CCAACAAAAGGCCTATCAAGAAATCAAGCAAGCTCTTCTAACTGC CCCAGCCCTGGGGTTGCCAGATTTGACTAAGCCCTTTGAACTCTTT GTCGACGAGAAGCAGGGCTACGCCAAAGGTGTCCTAACGCAAAA ACTGGGACCTTGGCGTCGGCCGGTGGCCTACCTGTCCAAAAAGCT AGACCCAGTAGCAGCTGGGTGGCCCCCTTGCCTACGGATGGTAGC AGCCATTGCCGTACTGACAAAGGATGCAGGCAAGCTAACCATGGG ACAGCCACTAGTCATTCTGGCCCCCCATGCAGTAGAGGCACTAGT CAAACAACCCCCCGACCGCTGGCTTTCCAACGCCCGGATGACTCA CTATCAGGCCTTGCTTTTGGACACGGACCGGGTCCAGTTCGGACCG GTGGTAGCCCTGAACCCGGCTACGCTGCTCCCACTGCCTGAGGAA GGGCTGCAACACAACTGCCTTGATATCCTGGCCGAAGCCCACGGA ACCCGACCCGACCTAACGGACCAGCCGCTCCCAGACGCCGACCAC ACCTGGTACACGGATGGAAGCAGTCTCTTACAAGAGGGACAGCGT AAGGCGGGAGCTGCGGTGACCACCGAGACCGAGGTAATCTGGGCT AAAGCCCTGCCAGCCGGGACATCCGCTCAGCGGGCTGAACTGATA GCACTCACCCAGGCCCTAAAGATGGCAGAAGGTAAGAAGCTAAAT GTTTATACTGATAGCCGTTATGCTTTTGCTACTGCCCATATCCATG GAGAAATATACAGAAGGCGTGGGTGGCTCACATCAGAAGGCAAA GAGATCAAAAATAAAGACGAGATCTTGGCCCTACTAAAAGCCCTC TTTCTGCCCAAAAGACTTAGCATAATCCATTGTCCAGGACATCAAA AGGGACACAGCGCCGAGGCTAGAGGCAACCGGATGGCTGACCAA GCGGCCCGAAAGGCAGCCATCACAGAGACTCCAGACACCTCTACC CTCCTCATAGAAAATTCATCACCCTCTGGCGGCTCAAAAAGAACC GCCGACGGCAGCGAATTCGAGCCCAAGAAGAAGAGGAAAGTCGG AAGCGGAGCTACTAACTTCAGCCTGCTGAAGCAGGCTGGCGACGT GGAGGAGAACCCTGGACCTAATTTACTGACCGTACACCAAAATTT GCCTGCATTACCGGTCGATGCAACGAGTGATGAGGTTCGCAAGAA CCTGATGGACATGTTCAGGGATCGCCAGGCGTTTTCTGAGCATACC TGGAAAATGCTTCTGTCCGTTTGCCGGTCGTGGGCGGCATGGTGCA AGTTGAATAACCGGAAATGGTTTCCCGCAGAACCTGAAGATGTTC GCGATTATCTTCTATATCTTCAGGCGCGCGGTCTGGCAGTAAAAAC TATCCAGCAACATTTGGGCCAGCTAAACATGCTTCATCGTCGGTCC GGGCTGCCACGACCAAGTGACAGCAATGCTGTTTCACTGGTTATG CGGCGGATCCGAAAAGAAAACGTTGATGCCGGTGAACGTGCAAA ACAGGCTCTAGCGTTCGAACGCACTGATTTCGACCAGGTTCGTTCA CTCATGGAAAATAGCGATCGCTGCCAGGATATACGTAATCTGGCA TTTCTGGGGATTGCTTATAACACCCTGTTACGTATAGCCGAAATTG CCAGGATCAGGGTTAAAGATATCTCACGTACTGACGGTGGGAGAA TGTTAATCCATATTGGCAGAACGAAAACGCTGGTTAGCACCGCAG GTGTAGAGAAGGCACTTAGCCTGGGGGTAACTAAACTGGTCGAGC GATGGATTTCCGTCTCTGGTGTAGCTGATGATCCGAATAACTACCT GTTTTGCCGGGTCAGAAAAAATGGTGTTGCCGCGCCATCTGCCAC CAGCCAGCTATCAACTCGCGCCCTGGAAGGGATTTTTGAAGCAAC TCATCGATTGATTTACGGCGCTAAGGATGACTCTGGTCAGAGATA CCTGGCCTGGTCTGGACACAGTGCCCGTGTCGGAGCCGCGCGAGA

	TATGGCCCGCGCTGGAGTTTCAATACCGGAGATCATGCAAGCTGG TGGCTGGACCAATGTAAATATTGTCATGAACTATATCCGTAACCTG GATAGTGAAACAGGGGCAATGGTGCGCCTGCTGGAAGATGGCGAT TAATTTAAACCCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCA GCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAA GGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGAAAATTGCAT CGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGG GCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATG CTGGGGATGCGGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACCA GCTGGGGCTCGATACCGTCGACCTCTAGCTAGAGCTTGGCGTAAT CATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAAT TCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTAGGG TGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTG CCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAA TCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTT CCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCG GCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCAC AGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCC AGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTT TCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCT CAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAG GCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCC TGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGT GGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAG GTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGC CCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCC GGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAG GATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAA GTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTAT CTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAG CTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTT GTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGA AGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAA AACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATC TTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCT AAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAAT CAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATA GTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGC TTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGC TCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGG GCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGT CTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTA ATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTC ACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGA TCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTT AGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAG TGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGT CATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACC AAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCC CGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAA AAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAA GGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGC ACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGG

	TGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAG GGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATAT TATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATAT TTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACAT TTCCCCGAAAAGTGCCACCTGACGTCGACGGATCGGGAGATCGAT CTCCCGATCCCCTAGGGTCGACTCTCAGTACAATCTGCTCTGATGC CGCATAGTTAAGCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTC GCTGAGTAGTGCGCGAGCAAAATTTAAGCTACAACAAGGCAAGGC TTGACCGACAATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTG CGCTGCTTCGCGATGTACGGGCCAGATAT
SEQ ID NO: 74 +90 ngRNA guide sequence (Artificial Sequence)	GTCAACCAGTATCCCGGTGC
SEQ ID NO: 75 +90 ngRNA (Artificial Sequence)	GTCAACCAGTATCCCGGTGCGTTTTAGAGCTAGAAATAGCAAGTT AAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGT CGGTGC
SEQ ID NO: 76 GFP minicircle template (before cleavage into a minicircle) (Artificial Sequence)	TGATCCCCTGCGCCATCAGATCCTTGGCGGCGAGAAAGCCATCCA GTTTACTTTGCAGGGCTTCCCAACCTTACCAGAGGGCGCCCCAGCT GGCAATTCCGGTTCGCTTGCTGTCCATAAAACCGCCCAGTCTAGCT ATCGCCATGTAAGCCCACTGCAAGCTACCTGCTTTCTCTTTGCGCT TGCGTTTTCCCTTGTCCAGATAGCCCAGTAGCTGACATTCATCCGG GGTCAGCACCGTTTCTGCGGACTGGCTTTCTACGTGCTCGAGGGGG GCCAAACGGTCTCCAGCTTGGCTGTTTTGGCGGATGAGAGAAGAT TTTCAGCCTGATACAGATTAAATCAGAACGCAGAAGCGGTCTGAT AAAACAGAATTTGCCTGGCGGCAGTAGCGCGGTGGTCCCACCTGA CCCCATGCCGAACTCAGAAGTGAAACGCCGTAGCGCCGATGGTAG TGTGGGGTCTCCCCATGCGAGAGTAGGGAACTGCCAGGCATCAAA TAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTTTTATCT GTTGTTTGTCGGTGAACGCTCTCCTGAGTAGGACAAATCCGCCGG GAGCGGATTTGAACGTTGCGAAGCAACGGCCCGGAGGGTGGCGG GCAGGACGCCCGCCATAAACTGCCAGGCATCAAATTAAGCAGAAG GCCATCCTGACGGATGGCCTTTTTGCGTTTCTACAAACTCTTTTGTT TATTTTTCTAAATACATTCAAATATGTATCCGCTCATGACCAAAAT CCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAA AAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCT GCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTT TGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTT CAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTA GTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTC GCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGT CGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGG CGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCT TGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGC TATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGG

	TATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGA GCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTT CGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGG GGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGT TCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTA TCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTG ATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGA GCGAGGAAGCGGAAGAGCGCCTGATGCGGTATTTTCTCCTTACGC ATCTGTGCGGTATTTCACACCGCATATGGTGCACTCTCAGTACAAT CTGCTCTGATGCCGCATAGTTAAGCCAGTATACACTCCGCTATCGC TACGTGACTGGGTCATGGCTGCGCCCCGACACCCGCCAACACCCG CTGACGCGCCCTGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAG ACAAGCTGTGACCGTCTCCGGGAGCTGCATGTGTCAGAGGTTTTC ACCGTCATCACCGAAACGCGCGAGGCAGCAGATCAATTCGCGCGC GAAGGCGAAGCGGCATGCATAATGTGCCTGTCAAATGGACGAAGC AGGGATTCTGCAAACCCTATGCTACTCCGTCAAGCCGTCAATTGTC TGATTCGTTACCAATTATGACAACTTGACGGCTACATCATTCACTT TTTCTTCACAACCGGCACGGAACTCGCTCGGGCTGGCCCCGGTGC ATTTTTTAAATACCCGCGAGAAATAGAGTTGATCGTCAAAACCAA CATTGCGACCGACGGTGGCGATAGGCATCCGGGTGGTGCTCAAAA GCAGCTTCGCCTGGCTGATACGTTGGTCCTCGCGCCAGCTTAAGAC GCTAATCCCTAACTGCTGGCGGAAAAGATGTGACAGACGCGACGG CGACAAGCAAACATGCTGTGCGACGCTGGCGATACATTACCCTGT TATCCCTAGATGACATTACCCTGTTATCCCAGATGACATTACCCTG TTATCCCTAGATGACATTACCCTGTTATCCCTAGATGACATTTACC CTGTTATCCCTAGATGACATTACCCTGTTATCCCAGATGACATTAC CCTGTTATCCCTAGATACATTACCCTGTTATCCCAGATGACATACC CTGTTATCCCTAGATGACATTACCCTGTTATCCCAGATGACATTAC CCTGTTATCCCTAGATACATTACCCTGTTATCCCAGATGACATACC CTGTTATCCCTAGATGACATTACCCTGTTATCCCAGATGACATTAC CCTGTTATCCCTAGATACATTACCCTGTTATCCCAGATGACATACC CTGTTATCCCTAGATGACATTACCCTGTTATCCCAGATGACATTAC CCTGTTATCCCTAGATACATTACCCTGTTATCCCAGATGACATACC CTGTTATCCCTAGATGACATTACCCTGTTATCCCAGATGACATTAC CCTGTTATCCCTAGATACATTACCCTGTTATCCCAGATGACATACC CTGTTATCCCTAGATGACATTACCCTGTTATCCCAGATAAACTCAA TGATGATGATGATGATGGTCGAGACTCAGCGGCCGCGGTGCCAGG GCGTGCCCTTGGGCTCCCCGGGCGCGACTATAAGCTGCGAGCAAC TTCACTTGGGTATGCCGGCGGTAGCGCTTACCGTTCGTATAATGTA TGCTATACGAAGTTATCCGAAGCCGCTAGCGGTGGTTTGTCTGGTC AACCACCGCGGTCTCAGTGGTGTACGGTACAAACCCAGCTACCGG TCGCCACCATGCCCGCCATGAAGATCGAGTGCCGCATCACCGGCA CCCTGAACGGCGTGGAGTTCGAGCTGGTGGGCGGCGGAGAGGGC ACCCCCGAGCAGGGCCGCATGACCAACAAGATGAAGAGCACCAA AGGCGCCCTGACCTTCAGCCCCTACCTGCTGAGCCACGTGATGGG CTACGGCTTCTACCACTTCGGCACCTACCCCAGCGGCTACGAGAA CCCCTTCCTGCACGCCATCAACAACGGCGGCTACACCAACACCCG CATCGAGAAGTACGAGGACGGCGGCGTGCTGCACGTGAGCTTCAG CTACCGCTACGAGGCCGGCCGCGTGATCGGCGACTTCAAGGTGGT GGGCACCGGCTTCCCCGAGGACAGCGTGATCTTCACCGACAAGAT CATCCGCAGCAACGCCACCGTGGAGCACCTGCACCCCATGGGCGA

	TAACGTGCTGGTGGGCAGCTTCGCCCGCACCTTCAGCCTGCGCGA CGGCGGCTACTACAGCTTCGTGGTGGACAGCCACATGCACTTCAA GAGCGCCATCCACCCCAGCATCCTGCAGAACGGGGGCCCCATGTT CGCCTTCCGCCGCGTGGAGGAGCTGCACAGCAACACCGAGCTGGG CATCGTGGAGTACCAGCACGCCTTCAAGACCCCCATCGCCTTCGCC AGATCTCGAGCTCGATGAGTTTGGACAAACCACAACTAGAATGCA GTGAAAAAAAATGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTA TTTGTGGGCCCGCCCCAACTGGGGTAACCTTTGAGTTCTCTCAGTT GGGGGTAATCAGCATCATGATGTGGTACCACATCATGATGCTGAT TATAAGAATGCGGCCGCCACACTCTAGTGGATCTCGAGTTAATAA TTCAGAAGAACTCGTCAAGAAGGCGATAGAAGGCGATGCGCTGCG AATCGGGAGCGGCGATACCGTAAAGCACGAGGAAGCGGTCAGCC CATTCGCCGCCAAGCTCTTCAGCAATATCACGGGTAGCCAACGCT ATGTCCTGATAGCGGTCCGCCACACCCAGCCGGCCACAGTCGATG AATCCAGAAAAGCGGCCATTTTCCACCATGATATTCGGCAAGCAG GCATCGCCATGGGTCACGACGAGATCCTCGCCGTCGGGCATGCTC GCCTTGAGCCTGGCGAACAGTTCGGCTGGCGCGAGCCCCTGATGC TCTTCGTCCAGATCATCCTGATCGACAAGACCGGCTTCCATCCGAG TACGTGCTCGCTCGATGCGATGTTTCGCTTGGTGGTCGAATGGGCA GGTAGCCGGATCAAGCGTATGCAGCCGCCGCATTGCATCAGCCAT GATGGATACTTTCTCGGCAGGAGCAAGGTGTAGATGACATGGAGA TCCTGCCCCGGCACTTCGCCCAATAGCAGCCAGTCCCTTCCCGCTT CAGTGACAACGTCGAGCACAGCTGCGCAAGGAACGCCCGTCGTGG CCAGCCACGATAGCCGCGCTGCCTCGTCTTGCAGTTCATTCAGGGC ACCGGACAGGTCGGTCTTGACAAAAAGAACCGGGCGCCCCTGCGC TGACAGCCGGAACACGGCGGCATCAGAGCAGCCGATTGTCTGTTG TGCCCAGTCATAGCCGAATAGCCTCTCCACCCAAGCGGCCGGAGA ACCTGCGTGCAATCCATCTTGTTCAATCATGCGAAACGATCCTCAT CCTGTCTCTTGATCAGAGCT
SEQ ID NO: 77 Gaussia Luciferase minicircle template (Artificial Sequence)	TGATCCCCTGCGCCATCAGATCCTTGGCGGCGAGAAAGCCATCCA GTTTACTTTGCAGGGCTTCCCAACCTTACCAGAGGGCGCCCCAGCT GGCAATTCCGGTTCGCTTGCTGTCCATAAAACCGCCCAGTCTAGCT ATCGCCATGTAAGCCCACTGCAAGCTACCTGCTTTCTCTTTGCGCT TGCGTTTTCCCTTGTCCAGATAGCCCAGTAGCTGACATTCATCCGG GGTCAGCACCGTTTCTGCGGACTGGCTTTCTACGTGCTCGAGGGGG GCCAAACGGTCTCCAGCTTGGCTGTTTTGGCGGATGAGAGAAGAT TTTCAGCCTGATACAGATTAAATCAGAACGCAGAAGCGGTCTGAT AAAACAGAATTTGCCTGGCGGCAGTAGCGCGGTGGTCCCACCTGA CCCCATGCCGAACTCAGAAGTGAAACGCCGTAGCGCCGATGGTAG TGTGGGGTCTCCCCATGCGAGAGTAGGGAACTGCCAGGCATCAAA TAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTTTTATCT GTTGTTTGTCGGTGAACGCTCTCCTGAGTAGGACAAATCCGCCGG GAGCGGATTTGAACGTTGCGAAGCAACGGCCCGGAGGGTGGCGG GCAGGACGCCCGCCATAAACTGCCAGGCATCAAATTAAGCAGAAG GCCATCCTGACGGATGGCCTTTTTGCGTTTCTACAAACTCTTTTGTT TATTTTTCTAAATACATTCAAATATGTATCCGCTCATGACCAAAAT CCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAA AAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCT GCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTT TGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTT CAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTA GTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTC

	GCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGT CGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGG CGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCT TGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGC TATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGG TATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGA GCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTT CGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGG GGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGT TCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTA TCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTG ATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGA GCGAGGAAGCGGAAGAGCGCCTGATGCGGTATTTTCTCCTTACGC ATCTGTGCGGTATTTCACACCGCATATGGTGCACTCTCAGTACAAT CTGCTCTGATGCCGCATAGTTAAGCCAGTATACACTCCGCTATCGC TACGTGACTGGGTCATGGCTGCGCCCCGACACCCGCCAACACCCG CTGACGCGCCCTGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAG ACAAGCTGTGACCGTCTCCGGGAGCTGCATGTGTCAGAGGTTTTC ACCGTCATCACCGAAACGCGCGAGGCAGCAGATCAATTCGCGCGC GAAGGCGAAGCGGCATGCATAATGTGCCTGTCAAATGGACGAAGC AGGGATTCTGCAAACCCTATGCTACTCCGTCAAGCCGTCAATTGTC TGATTCGTTACCAATTATGACAACTTGACGGCTACATCATTCACTT TTTCTTCACAACCGGCACGGAACTCGCTCGGGCTGGCCCCGGTGC ATTTTTTAAATACCCGCGAGAAATAGAGTTGATCGTCAAAACCAA CATTGCGACCGACGGTGGCGATAGGCATCCGGGTGGTGCTCAAAA GCAGCTTCGCCTGGCTGATACGTTGGTCCTCGCGCCAGCTTAAGAC GCTAATCCCTAACTGCTGGCGGAAAAGATGTGACAGACGCGACGG CGACAAGCAAACATGCTGTGCGACGCTGGCGATACATTACCCTGT TATCCCTAGATGACATTACCCTGTTATCCCAGATGACATTACCCTG TTATCCCTAGATGACATTACCCTGTTATCCCTAGATGACATTTACC CTGTTATCCCTAGATGACATTACCCTGTTATCCCAGATGACATTAC CCTGTTATCCCTAGATACATTACCCTGTTATCCCAGATGACATACC CTGTTATCCCTAGATGACATTACCCTGTTATCCCAGATGACATTAC CCTGTTATCCCTAGATACATTACCCTGTTATCCCAGATGACATACC CTGTTATCCCTAGATGACATTACCCTGTTATCCCAGATGACATTAC CCTGTTATCCCTAGATACATTACCCTGTTATCCCAGATGACATACC CTGTTATCCCTAGATGACATTACCCTGTTATCCCAGATGACATTAC CCTGTTATCCCTAGATACATTACCCTGTTATCCCAGATGACATACC CTGTTATCCCTAGATGACATTACCCTGTTATCCCAGATGACATTAC CCTGTTATCCCTAGATACATTACCCTGTTATCCCAGATGACATACC CTGTTATCCCTAGATGACATTACCCTGTTATCCCAGATAAACTCAA TGATGATGATGATGATGGTCGAGACTCAGCGGCCGCGGTGCCAGG GCGTGCCCTTGGGCTCCCCGGGCGCGACTATAAGCTGCGAGCAAC TTCACTTGGGTATGCCGGCGGTAGCGCTTACCGTTCGTATAATGTA TGCTATACGAAGTTATCCGAAGCCGCTAGCGGTGGTTTGTCTGGTC AACCACCGCGGTCTCAGTGGTGTACGGTACAAACCCACTACCGGT CGCCACCATGGGAGTCAAAGTTCTGTTTGCCCTGATCTGCATCGCT GTGGCCGAGGCCAAGCCCACCGAGAACAACGAAGACTTCAACATC GTGGCCGTGGCCAGCAACTTCGCGACCACGGATCTCGATGCTGAC CGCGGGAAGTTGCCCGGCAAGAAGCTGCCGCTGGAGGTGCTCAAA GAGATGGAAGCCAATGCCCGGAAAGCTGGCTGCACCAGGGGCTGT

	CTGATCTGCCTGTCCCACATCAAGTGCACGCCCAAGATGAAGAAG TTCATCCCAGGACGCTGCCACACCTACGAAGGCGACAAAGAGTCC GCACAGGGCGGCATAGGCGAGGCGATCGTCGACATTCCTGAGATT CCTGGGTTCAAGGACTTGGAGCCCATGGAGCAGTTCATCGCACAG GTCGATCTGTGTGTGGACTGCACAACTGGCTGCCTCAAAGGGCTT GCCAACGTGCAGTGTTCTGACCTGCTCAAGAAGTGGCTGCCGCAA CGCTGTGCGACCTTTGCCAGCAAGATCCAGGGCCAGGTGGACAAG ATCAAGGGGGCCGGTGGTGACTAAGCGGAGCTCGATGAGTTTGGA CAAACCACAACTAGAATGCAGTGAAAAAAATGCTTTATTTGTGAA ATTTGTGATGCTATTGCTTTATTTGTGGGCCCGCCCCAACTGGGGT AACCTTTGAGTTCTCTCAGTTGGGGGTAATCAGCATCATGATGTGG TACCACATCATGATGCTGATTATAAGAATGCGGCCGCCACACTCT AGTGGATCTCGAGTTAATAATTCAGAAGAACTCGTCAAGAAGGCG ATAGAAGGCGATGCGCTGCGAATCGGGAGCGGCGATACCGTAAA GCACGAGGAAGCGGTCAGCCCATTCGCCGCCAAGCTCTTCAGCAA TATCACGGGTAGCCAACGCTATGTCCTGATAGCGGTCCGCCACAC CCAGCCGGCCACAGTCGATGAATCCAGAAAAGCGGCCATTTTCCA CCATGATATTCGGCAAGCAGGCATCGCCATGGGTCACGACGAGAT CCTCGCCGTCGGGCATGCTCGCCTTGAGCCTGGCGAACAGTTCGG CTGGCGCGAGCCCCTGATGCTCTTCGTCCAGATCATCCTGATCGAC AAGACCGGCTTCCATCCGAGTACGTGCTCGCTCGATGCGATGTTTC GCTTGGTGGTCGAATGGGCAGGTAGCCGGATCAAGCGTATGCAGC CGCCGCATTGCATCAGCCATGATGGATACTTTCTCGGCAGGAGCA AGGTGTAGATGACATGGAGATCCTGCCCCGGCACTTCGCCCAATA GCAGCCAGTCCCTTCCCGCTTCAGTGACAACGTCGAGCACAGCTG CGCAAGGAACGCCCGTCGTGGCCAGCCACGATAGCCGCGCTGCCT CGTCTTGCAGTTCATTCAGGGCACCGGACAGGTCGGTCTTGACAA AAAGAACCGGGCGCCCCTGCGCTGACAGCCGGAACACGGCGGCA TCAGAGCAGCCGATTGTCTGTTGTGCCCAGTCATAGCCGAATAGC CTCTCCACCCAAGCGGCCGGAGAACCTGCGTGCAATCCATCTTGTT CAATCATGCGAAACGATCCTCATCCTGTCTCTTGATCAGAGCT
SEQ ID NO: 78 pseudo attP site (Artificial sequence)	CCCCAACTGGGGTAACCTTTGAGTTCTCTCAGTTGGGG
SEQ ID NO: 79 Albumin-pegRNA-SERPIN (Artificial Sequence)	GACTGAAACTTCACAGAATAGTTTTAGAGCTAGAAATAGCAAGTT AAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGT CGGTGCTTGGGATAGTTATGAATTCAATCTTCAACCCTATCCGGAT GATCCTGACGACGGAGACCGCCGTCGTCGACAAGCCGGCCTCTGT GAAGTTTCAGTCA
SEQ ID NO: 80 Albumin-pegRNA-CPS1 (Artificial Sequence)	GACTGAAACTTCACAGAATAGTTTTAGAGCTAGAAATAGCAAGTT AAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGT CGGTGCTTGGGATAGTTATGAATTCAATCTTCAACCCTATCCGGAT GATCCTGACGACGGAGACCGCCGTCGTCGACAAGCCGGCCTCTGT GAAGTTTC

SEQ ID NO: 81 34 bp lox71 pegRNA (Artificial Sequence)	GGCCCAGACTGAGCACGTGAGTTTTAGAGCTAGAAATAGCAAGTT AAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGT CGGTGCTGGAGGAAGCAGGGCTTCCTTTCCTCTGCCATCATACCGT TCGTATAGCATACATTATACGAAGTTATCGTGCTCAGTCTG
SEQ ID NO: 82 34 bp lox66 pegRNA (Artificial Sequence)	GGCCCAGACTGAGCACGTGAGTTTTAGAGCTAGAAATAGCAAGTT AAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGT CGGTGCTGGAGGAAGCAGGGCTTCCTTTCCTCTGCCATCAATAACT TCGTATAGCATACATTATACGAACGGTACGTGCTCAGTCTG
SEQ ID NO: 83 gRNA (Artificial Sequence)	GGCCCAGACTGAGCACGTGA
SEQ ID NO: 84 ACTB N-term PBS 13 RT 29 attB 46 (original length) pegRNA (Artificial Sequence)	GCTATTCTCGCAGCTCACCAGTTTTAGAGCTAGAAATAGCAAGTTA AAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTC GGTGCGACGAGCGCGGCGATATCATCATCCATGGCCGGATGATCC TGACGACGGAGACCGCCGTCGTCGACAAGCCGGCCTGAGCTGCGA GAA
SEQ ID NO: 85 ACTB N-term PBS_13_RT_29_with TP901-1 minimal attB f pegRNA (Artificial Sequence)	GCTATTCTCGCAGCTCACCAGTTTTAGAGCTAGAAATAGCAAGTT AAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAG TCGGTGCGAGTCGGTGCGACGAGCGCGGCGATATCATCATCCAT GGCACAATTAACATCTCAATCAAGGTAAATGCTTGAGCTGCGAG AA
SEQ ID NO: 86 ACTB N-term PBS_13_RT_29_with TP901-1 minimal attB rc pegRNA (Artificial Sequence)	GCTATTCTCGCAGCTCACCAGTTTTAGAGCTAGAAATAGCAAGTT AAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAG TCGGTGCGAGTCGGTGCGACGAGCGCGGCGATATCATCATCCAT GGAGCATTTACCTTGATTGAGATGTTAATTGTGTGAGCTGCGAGA A
SEQ ID NO: 87 ACTB N-term PBS_13_RT_29_with PhiBT1 minimal attB f pegRNA (Artificial Sequence)	GCTATTCTCGCAGCTCACCAGTTTTAGAGCTAGAAATAGCAAGTT AAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAG TCGGTGCGAGTCGGTGCGACGAGCGCGGCGATATCATCATCCAT GGCAGGTTTTTGACGAAAGTGATCCAGATGATCCAGTGAGCTGC GAGAA

SEQ ID NO: 88 ACTB N-term PBS_13_RT_29_with PhiBT1 minimal attB rc pegRNA (Artificial Sequence)	GCTATTCTCGCAGCTCACCAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGAGTCGGTGCGACGAGCGCGGCGATATCATCATCCATGGCTGGATCATCTGGATCACTTTCGTCAAAAACCTGTGAGCTGCGAGAA
SEQ ID NO: 89 ACTB N-term Nicking guide 1 +48 guide (Artificial Sequence)	GAAGCCGGCCTTGCACATGCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
SEQ ID NO: 90 ACTB N-term PBS_18_RT_16_with_L ox71_Cre pegRNA (Artificial Sequence)	GAAGCCGGCCTTGCACATGCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCATATCATCATCCATGGTACCGTTCGTATAGCATACATTATACGAAGTTATTGAGCTGCGAGAATAGCC
SEQ ID NO: 91 ACTB N-term PBS_13_RT_29_with_L ox71_Cre pegRNA (Artificial Sequence)	GAAGCCGGCCTTGCACATGCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGACGAGCGCGGCGATATCATCATCCATGGTACCGTTCGTATAGCATACATTATACGAAGTTATTGAGCTGCGAGAA
SEQ ID NO: 92 ACTB N-term PBS 13 RT 34 pegRNA (Artificial Sequence)	GCTATTCTCGCAGCTCACCAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTCGACGACGAGCGCGGCGATATCATCATCCATGGCCGGATGATCCTGACGACGGAGACCGCCGTCGTCGACAAGCCGGCCTGAGCTGCGAGAA
SEQ ID NO: 93 ACTB N-term PBS 13 RT 26 pegRNA (Artificial Sequence)	GCTATTCTCGCAGCTCACCAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGAGCGCGGCGATATCATCATCCATGGCCGGATGATCCTGACGACGGAGACCGCCGTCGTCGACAAGCCGGCCTGAGCTGCGAGAA

SEQ ID NO: 94 ACTB N-term PBS 13 RT 23 pegRNA (Artificial Sequence)	GCTATTCTCGCAGCTCACCAGTTTTAGAGCTAGAAATAGCAAGTTA AAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTC GGTGCCGCGGCGATATCATCATCCATGGCCGGATGATCCTGACGAC GGAGACCGCCGTCGTCGACAAGCCGGCCTGAGCTGCGAGAA
SEQ ID NO: 95 ACTB N-term PBS 13 RT 20 pegRNA (Artificial Sequence)	GCTATTCTCGCAGCTCACCAGTTTTAGAGCTAGAAATAGCAAGTTA AAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTC GGTGCGGCGATATCATCATCCATGGCCGGATGATCCTGACGACGG AGACCGCCGTCGTCGACAAGCCGGCCTGAGCTGCGAGAA
SEQ ID NO: 96 ACTB N-term PBS 13 RT 16 pegRNA (Artificial Sequence)	GCTATTCTCGCAGCTCACCAGTTTTAGAGCTAGAAATAGCAAGTTA AAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTC GGTGCATATCATCATCCATGGCCGGATGATCCTGACGACGGAGAC CGCCGTCGTCGACAAGCCGGCCTGAGCTGCGAGAA
SEQ ID NO: 97 ACTB N-term PBS 18 RT 34 pegRNA (Artificial Sequence)	GCTATTCTCGCAGCTCACCAGTTTTAGAGCTAGAAATAGCAAGTTA AAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTC GGTGCTCGACGACGAGCGCGGCGATATCATCATCCATGGCCGGAT GATCCTGACGACGGAGACCGCCGTCGTCGACAAGCCGGCCTGAGC TGCGAGAATAGCC
SEQ ID NO: 98 ACTB N-term PBS 18 RT 29 pegRNA (Artificial Sequence)	GCTATTCTCGCAGCTCACCAGTTTTAGAGCTAGAAATAGCAAGTTA AAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTC GGTGCGACGAGCGCGGCGATATCATCATCCATGGCCGGATGATCC TGACGACGGAGACCGCCGTCGTCGACAAGCCGGCCTGAGCTGCGA GAATAGCC
SEQ ID NO: 99 ACTB N-term PBS 18 RT 16 pegRNA (Artificial Sequence)	GCTATTCTCGCAGCTCACCAGTTTTAGAGCTAGAAATAGCAAGTTA AAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTC GGTGCATATCATCATCCATGGCCGGATGATCCTGACGACGGAGAC CGCCGTCGTCGACAAGCCGGCCTGAGCTGCGAGAATAGCC
SEQ ID NO: 100 LMNB1 N-term PBS 13 RT 39 pegRNA (Artificial Sequence)	GCTGTCTCCGCCGCCCGCCAGTTTTAGAGCTAGAAATAGCAAGTT AAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAG TCGGTGCCTGCCCATCCGCGGCGGCACGGGGGTCGCAGTCGCCA TGCCGGATGATCCTGACGACGGAGACCGCCGTCGTCGACAAGCC GGCCCGGGCGGCGGAGA

SEQ ID NO: 101 LMNB1 N-term PBS 13 RT 34 pegRNA (Artificial Sequence)	GCTGTCTCCGCCGCCCGCCAGTTTTAGAGCTAGAAATAGCAAGTT AAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAG TCGGTGCCATCCGCGGCGGCACGGGGGTCGCAGTCGCCATGCCG GATGATCCTGACGACGGAGACCGCCGTCGTCGACAAGCCGGCCC GGGCGGCGGAGA
SEQ ID NO: 102 LMNB1 N-term PBS 13 RT 29 pegRNA (Artificial Sequence)	GCTGTCTCCGCCGCCCGCCAGTTTTAGAGCTAGAAATAGCAAGTT AAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAG TCGGTGCGCGGCGGCACGGGGGTCGCAGTCGCCATGCCGGATGA TCCTGACGACGGAGACCGCCGTCGTCGACAAGCCGGCCCGGGCG GCGGAGA
SEQ ID NO: 103 LMNB1 N-term PBS 13 RT 24 pegRNA (Artificial Sequence)	GCTGTCTCCGCCGCCCGCCAGTTTTAGAGCTAGAAATAGCAAGTT AAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAG TCGGTGCGGCACGGGGGTCGCAGTCGCCATGCCGGATGATCCTG ACGACGGAGACCGCCGTCGTCGACAAGCCGGCCCGGGCGGCGGA GA
SEQ ID NO: 104 LMNB1 N-term PBS 13 RT 19 pegRNA (Artificial Sequence)	GCTGTCTCCGCCGCCCGCCAGTTTTAGAGCTAGAAATAGCAAGTT AAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAG TCGGTGCGGGGGTCGCAGTCGCCATGCCGGATGATCCTGACGAC GGAGACCGCCGTCGTCGACAAGCCGGCCCGGGCGGCGGAGA
SEQ ID NO: 105 LMNB1 N-term PBS 18 RT 39 pegRNA (Artificial Sequence)	GCTGTCTCCGCCGCCCGCCAGTTTTAGAGCTAGAAATAGCAAGTT AAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAG TCGGTGCCTGCCCATCCGCGGCGGCACGGGGGTCGCAGTCGCCA TGCCGGATGATCCTGACGACGGAGACCGCCGTCGTCGACAAGCC GGCCCGGGCGGCGGAGACAGCG
SEQ ID NO: 106 LMNB1 N-term PBS 18 RT 34 pegRNA (Artificial Sequence)	GCTGTCTCCGCCGCCCGCCAGTTTTAGAGCTAGAAATAGCAAGTT AAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAG TCGGTGCCATCCGCGGCGGCACGGGGGTCGCAGTCGCCATGCCG GATGATCCTGACGACGGAGACCGCCGTCGTCGACAAGCCGGCCC GGGCGGCGGAGACAGCG
SEQ ID NO: 107 LMNB1 N-term PBS 18 RT 29 pegRNA (Artificial Sequence)	GCTGTCTCCGCCGCCCGCCAGTTTTAGAGCTAGAAATAGCAAGTT AAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGT CGGTGCGCGGCGGCACGGGGGTCGCAGTCGCCATGCCGGATGATC CTGACGACGGAGACCGCCGTCGTCGACAAGCCGGCCCGGGCGGCG GAGACAGCG

SEQ ID NO: 108 LMNB1 N-term PBS 18 RT 24 pegRNA (Artificial Sequence)	GCTGTCTCCGCCGCCCGCCAGTTTTAGAGCTAGAAATAGCAAGTT AAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAG TCGGTGCGGCACGGGGGTCGCAGTCGCCATGCCGGATGATCCTG ACGACGGAGACCGCCGTCGTCGACAAGCCGGCCCGGGCGGCGGA GACAGCG
SEQ ID NO: 109 LMNB1 N-term PBS 18 RT 19 pegRNA (Artificial Sequence)	GCTGTCTCCGCCGCCCGCCAGTTTTAGAGCTAGAAATAGCAAGTT AAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAG TCGGTGCGGGGGTCGCAGTCGCCATGCCGGATGATCCTGACGAC GGAGACCGCCGTCGTCGACAAGCCGGCCCGGGCGGCGGAGACAG CG
SEQ ID NO: 110 LMNB1 N-term Nicking guide 1 +46 (Artificial Sequence)	GCGTGGTGGGGCCGCCAGCGGTTTTAGAGCTAGAAATAGCAAGT TAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGA GTCGGTGC
SEQ ID NO: 111 ACTB N-term PBS 13 RT 29 attB 42 pegRNA (Artificial Sequence)	GCTATTCTCGCAGCTCACCAGTTTTAGAGCTAGAAATAGCAAGTTA AAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTC GGTGCGACGAGCGCGGCGATATCATCATCCATGGGGATGATCCTG ACGACGGAGACCGCCGTCGTCGACAAGCCGGTGAGCTGCGAGAA
SEQ ID NO: 112 ACTB N-term PBS 13 RT 29 attB 40 pegRNA (Artificial Sequence)	GCTATTCTCGCAGCTCACCAGTTTTAGAGCTAGAAATAGCAAGTTA AAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTC GGTGCGACGAGCGCGGCGATATCATCATCCATGGGATGATCCTGA CGACGGAGACCGCCGTCGTCGACAAGCCGTGAGCTGCGAGAA
SEQ ID NO: 113 ACTB N-term PBS 13 RT 29 attB 38 pegRNA (Artificial Sequence)	GCTATTCTCGCAGCTCACCAGTTTTAGAGCTAGAAATAGCAAGTTA AAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTC GGTGCGACGAGCGCGGCGATATCATCATCCATGGATGATCCTGAC GACGGAGACCGCCGTCGTCGACAAGCCTGAGCTGCGAGAA
SEQ ID NO: 114 ACTB N-term PBS 13 RT 29 attB 36 pegRNA (Artificial Sequence)	GCTATTCTCGCAGCTCACCAGTTTTAGAGCTAGAAATAGCAAGTTA AAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTC GGTGCGACGAGCGCGGCGATATCATCATCCATGGTGATCCTGACG ACGGAGACCGCCGTCGTCGACAAGCTGAGCTGCGAGAA

SEQ ID NO: 115 LMNB1 N-term PBS 13 RT 29 attB 44 pegRNA v2 (Artificial Sequence)	GCTGTCTCCGCCGCCCGCCAGTTTTAGAGCTAGAAATAGCAAGTT AAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGT CGGTGCGCGGCGGCACGGGGGTCGCAGTCGCCATGCGGATGATCC TGACGACGGAGACCGCCGTCGTCGACAAGCCGGCCGGGCGGCGG AGA
SEQ ID NO: 116 LMNB1 N-term PBS 13 RT 29 attB 42 pegRNA v2 (Artificial Sequence)	GCTGTCTCCGCCGCCCGCCAGTTTTAGAGCTAGAAATAGCAAGTT AAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGT CGGTGCGCGGCGGCACGGGGGTCGCAGTCGCCATGGGATGATCCT GACGACGGAGACCGCCGTCGTCGACAAGCCGGCGGGCGGCGGAG A
SEQ ID NO: 117 LMNB1 N-term PBS 13 RT 29 attB 40 pegRNA v2 (Artificial Sequence)	GCTGTCTCCGCCGCCCGCCAGTTTTAGAGCTAGAAATAGCAAGTT AAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGT CGGTGCGCGGCGGCACGGGGGTCGCAGTCGCCATGGATGATCCTG ACGACGGAGACCGCCGTCGTCGACAAGCCGCGGGCGGCGGAGA
SEQ ID NO: 118 LMNB1 N-term PBS 13 RT 29 attB 38 pegRNA v2 (Artificial Sequence)	GCTGTCTCCGCCGCCCGCCAGTTTTAGAGCTAGAAATAGCAAGTT AAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGT CGGTGCGCGGCGGCACGGGGGTCGCAGTCGCCATGATGATCCTGA CGACGGAGACCGCCGTCGTCGACAAGCCCGGGCGGCGGAGA
SEQ ID NO: 119 NOLC1 N-term PBS 18 RT 29 attB 46 pegRNA (Artificial Sequence)	GCGTATTGCCTGGAGGATGGGTTTTAGAGCTAGAAATAGCAAGT TAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGA GTCGGTGCGAACCACGCGGCGAATGCCGGCGTCCGCCCCGGATG ATCCTGACGACGGAGACCGCCGTCGTCGACAAGCCGGCCTCCTC CAGGCAATACGCG
SEQ ID NO: 120 NOLC1 N-term PBS 13 RT 29 attB 46 pegRNA (Artificial Sequence)	GCGTATTGCCTGGAGGATGGGTTTTAGAGCTAGAAATAGCAAGTT AAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGT CGGTGCGAACCACGCGGCGAATGCCGGCGTCCGCCCCGGATGATC CTGACGACGGAGACCGCCGTCGTCGACAAGCCGGCCTCCTCCAGG CAAT

SEQ ID NO: 121 NOLC1 N-term PBS 13 RT 29 attB 44 pegRNA (Artificial Sequence)	GCGTATTGCCTGGAGGATGGGTTTTAGAGCTAGAAATAGCAAGT TAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGA GTCGGTGCGAACCACGCGGCGAATGCCGGCGTCCGCCCGGATGA TCCTGACGACGGAGACCGCCGTCGTCGACAAGCCGGCTCCTCCA GGCAAT
SEQ ID NO: 122 NOLC1 N-term PBS 13 RT 29 attB 42 pegRNA (Artificial Sequence)	GCGTATTGCCTGGAGGATGGGTTTTAGAGCTAGAAATAGCAAGT TAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGA GTCGGTGCGAACCACGCGGCGAATGCCGGCGTCCGCCGGATGAT CCTGACGACGGAGACCGCCGTCGTCGACAAGCCGGTCCTCCAGG CAAT
SEQ ID NO: 123 NOLC1 N-term PBS 13 RT 29 attB 40 pegRNA (Artificial Sequence)	GCGTATTGCCTGGAGGATGGGTTTTAGAGCTAGAAATAGCAAGT TAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGA GTCGGTGCGAACCACGCGGCGAATGCCGGCGTCCGCCGATGATC CTGACGACGGAGACCGCCGTCGTCGACAAGCCGTCCTCCAGGCA AT
SEQ ID NO: 124 NOLC1 N-term PBS 13 RT 29 attB 38 pegRNA (Artificial Sequence)	GCGTATTGCCTGGAGGATGGGTTTTAGAGCTAGAAATAGCAAGTT AAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAG TCGGTGCGAACCACGCGGCGAATGCCGGCGTCCGCCATGATCCT GACGACGGAGACCGCCGTCGTCGACAAGCCTCCTCCAGGCAAT
SEQ ID NO: 125 NOLC1 nicking guide - 43 (Artificial Sequence)	GAGCCGAGCACGAGGGGATACGTTTTAGAGCTAGAAATAGCAAGT TAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAG TCGGTGC
SEQ ID NO: 126 ACTB N-term PBS 13 RT 20 attB 38 pegRNA (Artificial Sequence)	GCTATTCTCGCAGCTCACCAGTTTTAGAGCTAGAAATAGCAAGTTA AAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTC GGTGCGGCGATATCATCATCCATGGATGATCCTGACGACGGAGAC CGCCGTCGTCGACAAGCCTGAGCTGCGAGAA
SEQ ID NO: 127 ACTB N-term PBS 13 RT 15 attB 38 pegRNA	GCTATTCTCGCAGCTCACCAGTTTTAGAGCTAGAAATAGCAAGTTA AAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTC GGTGCTATCATCATCCATGGATGATCCTGACGACGGAGACCGCCG TCGTCGACAAGCCTGAGCTGCGAGAA

(Artificial Sequence)	
SEQ ID NO: 128 ACTB N-term PBS 13 RT 10 attB 38 pegRNA (Artificial Sequence)	GCTATTCTCGCAGCTCACCAGTTTTAGAGCTAGAAATAGCAAGTTA AAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTC GGTGCTCATCCATGGATGATCCTGACGACGGAGACCGCCGTCGTC GACAAGCCTGAGCTGCGAGAA
SEQ ID NO: 129 ACTB N-term PBS 9 RT 20 attB 38 pegRNA (Artificial Sequence)	GCTATTCTCGCAGCTCACCAGTTTTAGAGCTAGAAATAGCAAGTT AAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAG TCGGTGCGGCGATATCATCATCCATGGATGATCCTGACGACGGAG ACCGCCGTCGTCGACAAGCCTGAGCTGCG
SEQ ID NO: 130 ACTB N-term PBS 9 RT 15 attB 38 pegRNA (Artificial Sequence)	GCTATTCTCGCAGCTCACCAGTTTTAGAGCTAGAAATAGCAAGTTA AAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTC GGTGCTATCATCATCCATGGATGATCCTGACGACGGAGACCGCCG TCGTCGACAAGCCTGAGCTGCG
SEQ ID NO: 131 ACTB N-term PBS 9 RT 10 attB 38 pegRNA (Artificial Sequence)	GCTATTCTCGCAGCTCACCAGTTTTAGAGCTAGAAATAGCAAGTTA AAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTC GGTGCTCATCCATGGATGATCCTGACGACGGAGACCGCCGTCGTC GACAAGCCTGAGCTGCG
SEQ ID NO: 132 LMNB1 N-term PBS 13 RT 20 attB 38 pegRNA (Artificial Sequence)	GCTGTCTCCGCCGCCCGCCAGTTTTAGAGCTAGAAATAGCAAGTT AAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAG TCGGTGCCGGGGGTCGCAGTCGCCATGATGATCCTGACGACGGA GACCGCCGTCGTCGACAAGCCCGGGCGGCGGAGA
SEQ ID NO: 133 LMNB1 N-term PBS 13 RT 15 attB 38 pegRNA	GCTGTCTCCGCCGCCCGCCAGTTTTAGAGCTAGAAATAGCAAGTT AAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAG TCGGTGCGTCGCAGTCGCCATGATGATCCTGACGACGGAGACCG CCGTCGTCGACAAGCCCGGGCGGCGGAGA

(Artificial Sequence)	
SEQ ID NO: 134 LMNB1 N-term PBS 13 RT 10 attB 38 pegRNA (Artificial Sequence)	GCTGTCTCCGCCGCCCGCCAGTTTTAGAGCTAGAAATAGCAAGTT AAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAG TCGGTGCAGTCGCCATGATGATCCTGACGACGGAGACCGCCGTC GTCGACAAGCCCGGGCGGCGGAGA
SEQ ID NO: 135 LMNB1 N-term PBS 9 RT 20 attB 38 pegRNA (Artificial Sequence)	GCTGTCTCCGCCGCCCGCCAGTTTTAGAGCTAGAAATAGCAAGTT AAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGT CGGTGCCGGGGGTCGCAGTCGCCATGATGATCCTGACGACGGAGA CCGCCGTCGTCGACAAGCCCGGGCGGCG
SEQ ID NO: 136 LMNB1 N-term PBS 9 RT 15 attB 38 pegRNA (Artificial Sequence)	GCTGTCTCCGCCGCCCGCCAGTTTTAGAGCTAGAAATAGCAAGTT AAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAG TCGGTGCGTCGCAGTCGCCATGATGATCCTGACGACGGAGACCG CCGTCGTCGACAAGCCCGGGCGGCG
SEQ ID NO: 137 LMNB1 N-term PBS 9 RT 10 attB 38 pegRNA (Artificial Sequence)	GCTGTCTCCGCCGCCCGCCAGTTTTAGAGCTAGAAATAGCAAGTT AAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGT CGGTGCAGTCGCCATGATGATCCTGACGACGGAGACCGCCGTCGT CGACAAGCCCGGGCGGCG
SEQ ID NO: 138 SUPT16H N-term PBS 13 RT 24 Bxb1-GT_Initial length (Artificial Sequence)	GAGAAGCGGCGTCCGGGGCTAGTTTTAGAGCTAGAAATAGCAAGT TAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAG TCGGTGCTCTTTGTCCAGAGTCACAGCCATACCGGATGATCCTGAC GACGGAGACCGCCGTCGTCGACAAGCCGGCCCCCCGGACGCCGC
SEQ ID NO: 139 SRRM2 N-term PBS 13 RT 24 Bxb1 Initial length	GGGCACGGGGCCATGTACAAGTTTTAGAGCTAGAAATAGCAAGT TAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGA GTCGGTGCGGCGTCGGCAGCCCGATCCCGTTGCCGGATGATCCT GACGACGGAGACCGCCGTCGTCGACAAGCCGGCCTACATGGCCC CGT

(Artificial Sequence)	
SEQ ID NO: 140 DEPDC4 N-term PBS 18 RT 24 Bxb1 Initial length (Artificial Sequence)	GTGTCAGGTGGGGCGGGGCTAGTTTTAGAGCTAGAAATAGCAAG TTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCG AGTCGGTGCGCTGGCTCCTCCCCTGGCACCATACCGGATGATCCT GACGACGGAGACCGCCGTCGTCGACAAGCCGGCCCCCCGCCCCA CCTGACAC
SEQ ID NO: 141 NES N-term PBS 13 RT 29 Bxb1 Initial length (Artificial Sequence)	GAGTGGGTCAGACGAGCAGGAGTTTTAGAGCTAGAAATAGCAAGT TAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAG TCGGTGCGATGGAGGGCTGCATGGGGGAGGAGTCGCCGGATGATC CTGACGACGGAGACCGCCGTCGTCGACAAGCCGGCCTGCTCGTCT GACC
SEQ ID NO: 142 SUPT16H nicking guide -53 (Artificial Sequence)	GCAGCCACCCGCTCTCGGCCCGTTTTAGAGCTAGAAATAGCAAG TTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCG AGTCGGTGC
SEQ ID NO: 143 SRRM2 N-term nicking guide 1 +87 (Artificial Sequence)	GTGTAGTCAGGCCGCTCACCCGTTTTAGAGCTAGAAATAGCAAG TTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCG AGTCGGTGC
SEQ ID NO: 144 DEPDC4 N-term Nicking guide 1 +59 (Artificial Sequence)	GCTGACAAGTCTACGGAACCTGTTTTAGAGCTAGAAATAGCAAG TTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCG AGTCGGTGC
SEQ ID NO: 145 NES N-term Nicking guide 2 +79 (Artificial Sequence)	GCTCCTCCAGCGCCTTGACCGTTTTAGAGCTAGAAATAGCAAGTTA AAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTC GGTGC

SEQ ID NO: 146 HITI_ACTB_guide (Artificial Sequence)	GCTATTCTCGCAGCTCACCA
SEQ ID NO: 147 HITI_SUPTH16_guide (Artificial Sequence)	AGAAGCGGCGTCCGGGGCTA
SEQ ID NO: 148 HITI_SRRM2_guide (Artificial Sequence)	GGGCACGGGGCCATGTACAA
SEQ ID NO: 149 HITI_NOLC1_guide (Artificial Sequence)	GCGTATTGCCTGGAGGATGG
SEQ ID NO: 150 HITI_DEPDC4_guide (Artificial Sequence)	TGTCAGGTGGGGCGGGGCTA
SEQ ID NO: 151 HITI_NES_guide (Artificial Sequence)	AGTGGGTCAGACGAGCAGGA
SEQ ID NO: 152 HITI_LMNB1_guide (Artificial Sequence)	GCTGTCTCCGCCGCCCGCCA

SEQ ID NO: 153 HDR Cas9 ACTB guide (Artificial Sequence)	GCTATTCTCGCAGCTCACCAGTTTTAGAGCTAGAAATAGCAAGTT AAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAG TCGGTGC
SEQ ID NO: 154 ACTB N-term PBS 13 RT 29 attB original length pegRNAs for dinucleotides (Artificial Sequence)	GCTATTCTCGCAGCTCACCAGTTTTAGAGCTAGAAATAGCAAGTTA AAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTC GGTGCGACGAGCGCGGCGATATCATCATCCATGGCCGGATGATCC TGACGACGGAG**XX**CGCCGTCGTCGACAAGCCGGCCTGAGCTGCGA GAA **XX**: CG, GC, AT, TA, GG, TT, GA, AG, CC, TC, CT, AA, TG, GT, CA, or AC
SEQ ID NO: 155 ACTB N-term PBS 13 RT 29 pegRNA with attB 46 GT for fusion (Artificial Sequence)	GCTATTCTCGCAGCTCACCAGTTTTAGAGCTAGAAATAGCAAGTTA AAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTC GGTGCGACGAGCGCGGCGATATCATCATCCATGCCGGATGATCCT GACGACGGAGACCGCCGTCGTCGACAAGCCGGCCTGAGCTGCGAG AA
SEQ ID NO: 156 ACTB N-term PBS 13 RT 29 pegRNA with attB 46 CT for multiplexing (Artificial Sequence)	GCTATTCTCGCAGCTCACCAGTTTTAGAGCTAGAAATAGCAAGTTA AAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTC GGTGCGACGAGCGCGGCGATATCATCATCCATGCCGGATGATCCT GACGACGGAGAGCGCCGTCGTCGACAAGCCGGCCTGAGCTGCGA GAA
SEQ ID NO: 157 NOLC1 N-term PBS 18 RT 29 pegRNA with attB 46 GA for multiplexing (Artificial Sequence)	GCGTATTGCCTGGAGGATGGGTTTTAGAGCTAGAAATAGCAAGTT AAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGT CGGTGCGAACCACGCGGCGAATGCCGGCGTCCGCCCCGGATGATC CTGACGACGGAGTCCGCCGTCGTCGACAAGCCGGCCTCCTCCAGG CAATACGCG
SEQ ID NO: 158 LMNB1 N-term PBS 18 RT 29 pegRNA with attB 46 AG for multiplexing (Artificial Sequence)	GCTGTCTCCGCCGCCCGCCAGTTTTAGAGCTAGAAATAGCAAGTT AAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGT CGGTGCGCGGCGGCACGGGGGTCGCAGTCGCCATGCCGGATGATC CTGACGACGGAGCTCGCCGTCGTCGACAAGCCGGCCCGGGCGGCG GAGACAGCG

SEQ ID NO: 159 EMX1 Cas9 guide 1 (Artificial Sequence)	GTCACCTCCAATGACTAGGG
SEQ ID NO: 160 EMX1 Cas9 guide 2 (Artificial Sequence)	GGGCAACCACAAACCCACGA
SEQ ID NO: 161 ACTB N-term PBS 13 RT 29 attB 56 GA pegRNA (Artificial Sequence)	GCTATTCTCGCAGCTCACCAGTTTTAGAGCTAGAAATAGCAAGTTA AAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTC GGTGCGACGAGCGCGGCGATATCATCATCCATGGCTATGCCGGAT GATCCTGACGACGGAGTCCGCCGTCGTCGACAAGCCGGCCCTAGC TGAGCTGCGAGAA
SEQ ID NO: 162 ACTB N-term PBS 13 RT 29 attB 51 GA pegRNA (Artificial Sequence)	GCTATTCTCGCAGCTCACCAGTTTTAGAGCTAGAAATAGCAAGTTA AAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTC GGTGCGACGAGCGCGGCGATATCATCATCCATGGTGCCGGATGAT CCTGACGACGGAGTCCGCCGTCGTCGACAAGCCGGCCCTATGAGC TGCGAGAA
SEQ ID NO: 163 ACTB N-term PBS 13 RT 29 attB 46 GA pegRNA (Artificial Sequence)	GCTATTCTCGCAGCTCACCAGTTTTAGAGCTAGAAATAGCAAGTTA AAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTC GGTGCGACGAGCGCGGCGATATCATCATCCATGGCCGGATGATCC TGACGACGGAGTCCGCCGTCGTCGACAAGCCGGCCTGAGCTGCGA GAA
SEQ ID NO: 164 ACTB N-term PBS 13 RT 29 attB 41 GA pegRNA (Artificial Sequence)	GCTATTCTCGCAGCTCACCAGTTTTAGAGCTAGAAATAGCAAGTTA AAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTC GGTGCGACGAGCGCGGCGATATCATCATCCATGGGGATGATCCTG ACGACGGAGTCCGCCGTCGTCGACAAGCCGTGAGCTGCGAGAA
SEQ ID NO: 165 ACTB N-term PBS 13 RT 29 attB 36 GA pegRNA (Artificial Sequence)	GCTATTCTCGCAGCTCACCAGTTTTAGAGCTAGAAATAGCAAGTTA AAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTC GGTGCGACGAGCGCGGCGATATCATCATCCATGGTGATCCTGACG ACGGAGTCCGCCGTCGTCGACAAGCTGAGCTGCGAGAA

SEQ ID NO: 166 ACTB N-term PBS 13 RT 29 attB 31 GA pegRNA (Artificial Sequence)	GCTATTCTCGCAGCTCACCAGTTTTAGAGCTAGAAATAGCAAGTTA AAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTC GGTGCGACGAGCGCGGCGATATCATCATCCATGGATCCTGACGAC GGAGTCCGCCGTCGTCGACATGAGCTGCGAGAA
SEQ ID NO: 167 ACTB N-term PBS 13 RT 29 attB 26 GA pegRNA (Artificial Sequence)	GCTATTCTCGCAGCTCACCAGTTTTAGAGCTAGAAATAGCAAGTTA AAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTC GGTGCGACGAGCGCGGCGATATCATCATCCATGGCCTGACGACGG AGTCCGCCGTCGTCGTGAGCTGCGAGAA
SEQ ID NO: 168 ACTB N-term PBS 13 RT 29 attB 21 GA pegRNA (Artificial Sequence)	GCTATTCTCGCAGCTCACCAGTTTTAGAGCTAGAAATAGCAAGTTA AAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTC GGTGCGACGAGCGCGGCGATATCATCATCCATGGTGACGACGGAG TCCGCCGTCGTGAGCTGCGAGAA
SEQ ID NO: 169 ACTB N-term PBS 13 RT 29 attB 16 GA pegRNA (Artificial Sequence)	GCTATTCTCGCAGCTCACCAGTTTTAGAGCTAGAAATAGCAAGTTA AAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTC GGTGCGACGAGCGCGGCGATATCATCATCCATGGACGACGGAGTC CGCCGTGAGCTGCGAGAA
SEQ ID NO: 170 ACTB N-term PBS 13 RT 29 attB 11 GA pegRNA (Artificial Sequence)	GCTATTCTCGCAGCTCACCAGTTTTAGAGCTAGAAATAGCAAGTTA AAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTC GGTGCGACGAGCGCGGCGATATCATCATCCATGGGACGGAGTCCG TGAGCTGCGAGAA
SEQ ID NO: 171 ACTB N-term PBS 13 RT 29 attB 6 GA pegRNA (Artificial Sequence)	GCTATTCTCGCAGCTCACCAGTTTTAGAGCTAGAAATAGCAAGTTA AAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTC GGTGCGACGAGCGCGGCGATATCATCATCCATGGCGGAGTTGAGC TGCGAGAA

SEQ ID NO: 172 ACTB N-term PBS_18_RT_34_with_L ox71_Cre pegRNA (Artificial Sequence)	GAAGCCGGCCTTGCACATGCGTTTTAGAGCTAGAAATAGCAAGTT AAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGT CGGTGCTCGACGACGAGCGCGGCGATATCATCATCCATGGTACCG TTCGTATAGCATACATTATACGAAGTTATTGAGCTGCGAGAATAG CC
SEQ ID NO: 173 ACTB N-term PBS_18_RT_29_with_L ox71_Cre pegRNA (Artificial Sequence)	GAAGCCGGCCTTGCACATGCGTTTTAGAGCTAGAAATAGCAAGTT AAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGT CGGTGCGACGAGCGCGGCGATATCATCATCCATGGTACCGTTCGT ATAGCATACATTATACGAAGTTATTGAGCTGCGAGAATAGCC
SEQ ID NO: 174 ACTB N-term PBS_13_RT_34_with_L ox71_Cre pegRNA (Artificial Sequence)	GAAGCCGGCCTTGCACATGCGTTTTAGAGCTAGAAATAGCAAGTT AAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGT CGGTGCTCGACGACGAGCGCGGCGATATCATCATCCATGGTACCG TTCGTATAGCATACATTATACGAAGTTATTGAGCTGCGAGAA
SEQ ID NO: 175 ACTB N-term PBS_13_RT_16_with_L ox71_Cre pegRNA (Artificial Sequence)	GAAGCCGGCCTTGCACATGCGTTTTAGAGCTAGAAATAGCAAGTT AAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGT CGGTGCATATCATCATCCATGGTACCGTTCGTATAGCATACATTAT ACGAAGTTATTGAGCTGCGAGAA
SEQ ID NO: 176 ACTB N-term Nicking guide 2 +93 guide (Artificial Sequence)	CCCCACGATGGAGGGGAAGAGTTTTAGAGCTAGAAATAGCAAGTT AAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGT CGGTGC
SEQ ID NO: 177 LMNB1 N-term Nicking guide 2 +87 guide (Artificial Sequence)	CCTTCTCCTGGAGCCGCGACGTTTTAGAGCTAGAAATAGCAAGTT AAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGT CGGTGC

[310] Sequences of insertion sites can be found in Table 4 below.

Table 4

DESCRIPTION/ SOURCE	FORWARD SEQUENCE (5'-3')		REVERSE SEQUENCE (5'-3')	
	SEQ ID NO	Sequence	SEQ ID NO	Sequence
Bxb1_attP_GT_original _site (Artificial Sequence)	178	GTGGTTTGTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCCA	179	TGGGTTTGTACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCAC
Bxb1_attP_CG_site (Artificial Sequence)	180	GTGGTTTGTCTGGTCAACCACCGCGCGCTCAGTGGTGTACGGTACAAACCCA	181	TGGGTTTGTACCGTACACCACTGAGCGCGCGGTGGTTGACCAGACAAACCAC
Bxb1_attP_GC_site (Artificial Sequence)	182	GTGGTTTGTCTGGTCAACCACCGCGGCCTCAGTGGTGTACGGTACAAACCCA	183	TGGGTTTGTACCGTACACCACTGAGGCCGCGGTGGTTGACCAGACAAACCAC
Bxb1_attP_AT_site (Artificial Sequence)	184	GTGGTTTGTCTGGTCAACCACCGCGATCTCAGTGGTGTACGGTACAAACCCA	185	TGGGTTTGTACCGTACACCACTGAGATCGCGGTGGTTGACCAGACAAACCAC
Bxb1_attP_TA_site (Artificial Sequence)	186	GTGGTTTGTCTGGTCAACCACCGCGTACTCAGTGGTGTACGGTACAAACCCA	187	TGGGTTTGTACCGTACACCACTGAGTACGCGGTGGTTGACCAGACAAACCAC

Bxb1_attP_GG_site (Artificial Sequence)	188	GTGGTTTGTCTGGTCAACCACCGCGGGCTCAGTGGTGTACGGTACAAACCCA	189	TGGGTTTGTACCGTACACCACTGAGCCCGCGGTGGTTGACCAGACAAACCAC
Bxb1_attP_TT_site (Artificial Sequence)	190	GTGGTTTGTCTGGTCAACCACCGCGTTCTCAGTGGTGTACGGTACAAACCCA	191	TGGGTTTGTACCGTACACCACTGAGAACGCGGTGGTTGACCAGACAAACCAC
Bxb1_attP_GA_site (Artificial Sequence)	192	GTGGTTTGTCTGGTCAACCACCGCGGACTCAGTGGTGTACGGTACAAACCCA	193	TGGGTTTGTACCGTACACCACTGAGTCCGCGGTGGTTGACCAGACAAACCAC
Bxb1_attP_AG_site (Artificial Sequence)	194	GTGGTTTGTCTGGTCAACCACCGCGAGCTCAGTGGTGTACGGTACAAACCCA	195	TGGGTTTGTACCGTACACCACTGAGCTCGCGGTGGTTGACCAGACAAACCAC
Bxb1_attP_CC_site (Artificial Sequence)	196	GTGGTTTGTCTGGTCAACCACCGCGCCCTCAGTGGTGTACGGTACAAACCCA	197	TGGGTTTGTACCGTACACCACTGAGGGCGCGGTGGTTGACCAGACAAACCAC
Bxb1_attP_TC_site (Artificial Sequence)	198	GTGGTTTGTCTGGTCAACCACCGCGTCCTCAGTGGTGTACGGTACAAACCCA	199	TGGGTTTGTACCGTACACCACTGAGGACGCGGTGGTTGACCAGACAAACCAC
Bxb1_attP_CT_site (Artificial Sequence)	200	GTGGTTTGTCTGGTCAACCACCGCGCTCTCAGTGGTGTACGGTACAAACCCA	201	TGGGTTTGTACCGTACACCACTGAGAGCGCGGTGGTTGACCAGACAAACCAC

Bxb1_attP_AA_site (Artificial Sequence)	202	GTGGTTTGTCTGGTCAACCACCGCGAACTCAGTGGTGTACGGTACAAACCCA	203	TGGGTTTGTACCGTACACCACTGAGTTCGCGGTGGTTGACCAGACAAACCAC
Bxb1_attP_CA_site (Artificial Sequence)	204	GTGGTTTGTCTGGTCAACCACCGCGCACTCAGTGGTGTACGGTACAAACCCA	205	TGGGTTTGTACCGTACACCACTGAGTGCGCGGTGGTTGACCAGACAAACCAC
Bxb1_attP_AC_site (Artificial Sequence)	206	GTGGTTTGTCTGGTCAACCACCGCGACCTCAGTGGTGTACGGTACAAACCCA	207	TGGGTTTGTACCGTACACCACTGAGGTCGCGGTGGTTGACCAGACAAACCAC
Bxb1_attP_TG_site (Artificial Sequence)	208	GTGGTTTGTCTGGTCAACCACCGCGTGCTCAGTGGTGTACGGTACAAACCCA	209	TGGGTTTGTACCGTACACCACTGAGCACGCGGTGGTTGACCAGACAAACCAC
Bxb1_attB_46_GT_original_site (Artificial Sequence)	210	GGCCGGCTTGTCGACGACGGCGGTCTCCGTCGTCAGGATCATCCGG	211	CCGGATGATCCTGACGACGGAGACCGCCGTCGTCGACAAGCCGGCC
Bxb1_attB_46_AA_site (Artificial Sequence)	212	GGCCGGCTTGTCGACGACGGCGAACTCCGTCGTCAGGATCATCCGG	213	CCGGATGATCCTGACGACGGAGTTCGCCGTCGTCGACAAGCCGGCC
Bxb1_attB_46_GA_site (Artificial Sequence)	214	GGCCGGCTTGTCGACGACGGCGGACTCCGTCGTCAGGATCATCCGG	215	CCGGATGATCCTGACGACGGAGTCCGCCGTCGTCGACAAGCCGGCC

Bxb1_attB_46_CA_site (Artificial Sequence)	216	GGCCGGCTTGTCGACGACGGCGCACTCCGTCGTCAGGATCATCCGG	217	CCGGATGATCCTGACGACGGAGTGCGCCGTCGTCGACAAGCCGGCC
Bxb1_attB_46_TA_site (Artificial Sequence)	218	GGCCGGCTTGTCGACGACGGCGTACTCCGTCGTCAGGATCATCCGG	219	CCGGATGATCCTGACGACGGAGTACGCCGTCGTCGACAAGCCGGCC
Bxb1_attB_46_AG_site (Artificial Sequence)	220	GGCCGGCTTGTCGACGACGGCGAGCTCCGTCGTCAGGATCATCCGG	221	CCGGATGATCCTGACGACGGAGCTCGCCGTCGTCGACAAGCCGGCC
Bxb1_attB_46_GG_site (Artificial Sequence)	222	GGCCGGCTTGTCGACGACGGCGGGCTCCGTCGTCAGGATCATCCGG	223	CCGGATGATCCTGACGACGGAGCCCGCCGTCGTCGACAAGCCGGCC
Bxb1_attB_46_CG_site (Artificial Sequence)	224	GGCCGGCTTGTCGACGACGGCGCGCTCCGTCGTCAGGATCATCCGG	225	CCGGATGATCCTGACGACGGAGCGCGCCGTCGTCGACAAGCCGGCC
Bxb1_attB_46_TG_site (Artificial Sequence)	226	GGCCGGCTTGTCGACGACGGCGTGCTCCGTCGTCAGGATCATCCGG	227	CCGGATGATCCTGACGACGGAGCACGCCGTCGTCGACAAGCCGGCC
Bxb1_attB_46_AC_site (Artificial Sequence)	228	GGCCGGCTTGTCGACGACGGCGACCTCCGTCGTCAGGATCATCCGG	229	CCGGATGATCCTGACGACGGAGGTCGCCGTCGTCGACAAGCCGGCC

Bxb1_attB_46_GC_site (Artificial Sequence)	230	GGCCGGCTTGTCGACGACGGCGGCCTCCGTCGTCAGGATCATCCGG	231	CCGGATGATCCTGACGACGGAGGCCGCCGTCGTCGACAAGCCGGCC
Bxb1_attB_46_CC_site (Artificial Sequence)	232	GGCCGGCTTGTCGACGACGGCGCCCTCCGTCGTCAGGATCATCCGG	233	CCGGATGATCCTGACGACGGAGGGCGCCGTCGTCGACAAGCCGGCC
Bxb1_attB_46_TC_site (Artificial Sequence)	234	GGCCGGCTTGTCGACGACGGCGTCCTCCGTCGTCAGGATCATCCGG	235	CCGGATGATCCTGACGACGGAGGACGCCGTCGTCGACAAGCCGGCC
Bxb1_attB_46_AT_site (Artificial Sequence)	236	GGCCGGCTTGTCGACGACGGCGATCTCCGTCGTCAGGATCATCCGG	237	CCGGATGATCCTGACGACGGAGATCGCCGTCGTCGACAAGCCGGCC
Bxb1_attB_46_CT_site (Artificial Sequence)	238	GGCCGGCTTGTCGACGACGGCGCTCTCCGTCGTCAGGATCATCCGG	239	CCGGATGATCCTGACGACGGAGAGCGCCGTCGTCGACAAGCCGGCC
Bxb1_attB_46_TT_site (Artificial Sequence)	240	GGCCGGCTTGTCGACGACGGCGTTCTCCGTCGTCAGGATCATCCGG	241	CCGGATGATCCTGACGACGGAGAACGCCGTCGTCGACAAGCCGGCC
Bxb1_attB_38_GT_site (Artificial Sequence)	242	GGCTTGTCGACGACGGCGGTCTCCGTCGTCAGGATCAT	243	ATGATCCTGACGACGGAGACCGCCGTCGTCGACAAGCC

Bxb1_attB_38_AA_site (Artificial Sequence)	244	GGCTTGTCGACGACGGCGAACTCCGTCGTCAGGATCAT	245	ATGATCCTGACGACGGAGTTCGCCGTCGTCGACAAGCC
Bxb1_attB_38_GA_site (Artificial Sequence)	246	GGCTTGTCGACGACGGCGGACTCCGTCGTCAGGATCAT	247	ATGATCCTGACGACGGAGTCCGCCGTCGTCGACAAGCC
Bxb1_attB_38_CA_site (Artificial Sequence)	248	GGCTTGTCGACGACGGCGCACTCCGTCGTCAGGATCAT	249	ATGATCCTGACGACGGAGTGCGCCGTCGTCGACAAGCC
Bxb1_attB_38_TA_site (Artificial Sequence)	250	GGCTTGTCGACGACGGCGTACTCCGTCGTCAGGATCAT	251	ATGATCCTGACGACGGAGTACGCCGTCGTCGACAAGCC
Bxb1_attB_38_AG_site (Artificial Sequence)	252	GGCTTGTCGACGACGGCGAGCTCCGTCGTCAGGATCAT	253	ATGATCCTGACGACGGAGCTCGCCGTCGTCGACAAGCC
Bxb1_attB_38_GG_site (Artificial Sequence)	254	GGCTTGTCGACGACGGCGGGCTCCGTCGTCAGGATCAT	255	ATGATCCTGACGACGGAGCCCGCCGTCGTCGACAAGCC
Bxb1_attB_38_CG_site (Artificial Sequence)	256	GGCTTGTCGACGACGGCGCGCTCCGTCGTCAGGATCAT	257	ATGATCCTGACGACGGAGCGCGCCGTCGTCGACAAGCC

Bxb1_attB_38_TG_site (Artificial Sequence)	258	GGCTTGTCGACGAC GGCGTGCTCCGTCGT CAGGATCAT	259	ATGATCCTGACGAC GGAGCACGCCGTCG TCGACAAGCC
Bxb1_attB_38_AC_site (Artificial Sequence)	260	GGCTTGTCGACGAC GGCGACCTCCGTCGT CAGGATCAT	261	ATGATCCTGACGAC GGAGGTCGCCGTCG TCGACAAGCC
Bxb1_attB_38_GC_site (Artificial Sequence)	262	GGCTTGTCGACGAC GGCGGCCTCCGTCGT CAGGATCAT	263	ATGATCCTGACGAC GGAGGCCGCCGTCG TCGACAAGCC
Bxb1_attB_38_CC_site (Artificial Sequence)	264	GGCTTGTCGACGAC GGCGCCCTCCGTCGT CAGGATCAT	265	ATGATCCTGACGAC GGAGGGCGCCGTCG TCGACAAGCC
Bxb1_attB_38_TC_site (Artificial Sequence)	266	GGCTTGTCGACGAC GGCGTCCTCCGTCGT CAGGATCAT	267	ATGATCCTGACGAC GGAGGACGCCGTCG TCGACAAGCC
Bxb1_attB_38_AT_site (Artificial Sequence)	268	GGCTTGTCGACGAC GGCGATCTCCGTCGT CAGGATCAT	269	ATGATCCTGACGAC GGAGATCGCCGTCG TCGACAAGCC
Bxb1_attB_38_CT_site (Artificial Sequence)	270	GGCTTGTCGACGAC GGCGCTCTCCGTCGT CAGGATCAT	271	ATGATCCTGACGAC GGAGAGCGCCGTCG TCGACAAGCC

Bxb1_attB_38_TT_site (Artificial Sequence)	272	GGCTTGTCGACGAC GGCGTTCTCCGTCGT CAGGATCAT	273	ATGATCCTGACGAC GGAGAACGCCGTCG TCGACAAGCC
Cre Lox 66 site (Artificial Sequence)	274	TACCGTTCGTATAAT GTATGCTATACGAA GTTAT	275	ATAACTTCGTATAGC ATACATTATACGAA CGGTA
Cre Lox 71 site (Artificial Sequence)	276	ATAACTTCGTATAAT GTATGCTATACGAA CGGTA	277	TACCGTTCGTATAGC ATACATTATACGAA GTTAT
TP901-1 minimal attB site (Artificial Sequence)	278	TTTACCTTGATTGAG ATGTTAATTGTG	279	CACAATTAACATCTC AATCAAGGTAAA
TP901-1 minimal attP site (Artificial Sequence)	280	GCGAGTTTTTATTTC GTTTATTTCAATTAA GGTAACTAAAAAAC TCCTTT	281	AAAGGAGTTTTTTAG TTACCTTAATTGAAA TAAACGAAATAAAA ACTCGC
PhiBT1 minimal attB site (Artificial Sequence)	282	CTGGATCATCTGGAT CACTTTCGTCAAAAA CCTG	283	CAGGTTTTTGACGAA AGTGATCCAGATGA TCCAG
PhiBT1 minimal attP site (Artificial Sequence)	284	TTCGGGTGCTGGGTT GTTGTCTCTGGACAG TGATCCATGGGAAA CTACTCAGCACCA	285	TGGTGCTGAGTAGTT TCCCATGGATCACTG TCCAGAGACAACAA CCCAGCACCCGAA

[311] Sequences of Bxb1 and RT mutants can be found in Table 6 below.

Table 6	
SEQ ID NO/ DESCRIPTION/ SOURCE	**FORWARD SEQUENCE (5'-3')**
SEQ ID NO: 286 Bxb1_mut_V368A (Artificial Sequence)	AAAAGTGTGGGCTGCAGGATCTGA
SEQ ID NO: 287 Bxb1_mut_E379A (Artificial Sequence)	GGAGCTGGCAGCTGTCAATGCC
SEQ ID NO: 288 Bxb1_mut_E383A (Artificial Sequence)	AGTCAATGCCGCTCTCGTGGA
SEQ ID NO: 403 RT_mut_L139P (Artificial Sequence)	TTGAGCGGGCCCCCACCGT
SEQ ID NO: 289 RT_mut_E562Q (Artificial Sequence)	CAGCGGGCTCAGCTGATAGCA
SEQ ID NO: 290 RT_mut_D653N (Artificial Sequence)	CGGATGGCTAACCAAGCGGCC

SEQ ID NO: 404 RT(1-478)_Sto7d fusion	atgactcactatcaggccttgcttttggacacggaccgggtccagttcggaccggtggtagccctgaaccc ggctacgctgctcccactgcctgaggaagggctgcaacacaactgccttgatGGGACAGGTGG CGGTGGTGTCACCGTCAAGTTCAAGTACAAGGGTGAGGAACTT GAAGTTGATATTAGCAAAATCAAGAAGGTTTGGCGCGTTGGTA AAATGATATCTTTTACTTATGACGACAACGGCAAGACAGGTAG AGGGGCAGTGTCTGAGAAAGACGCCCCCAAGGAGCTGTTGCAA ATGTTGGAAAAGTCTGGGAAAAAGtctggcggctcaaaaagaaccgccgacgg cagcgaattcgagcccaagaagaagaggaaagtc

[312] Sequences of primers, probes and restriction enzymes used in ddPCR readout can be found in Table 7 below.

Table 7

Locus	**Cargo**	**SEQ ID NO:**	**Forward Primer**	**SEQ ID NO:**	**Reverse Primer**	**Probe**	**SEQ ID NO:**	**Restriction Enzymes**
ACTB	GFP (pDY0186)	291	CCCG GCTTC CTTTG TCC	292	GAAC TCCAC GCCG TTCA	/56-FAM/C C GGC TTG T/ZEN/ C GAC GAC GGC G/3IABkFQ/	405	Eco91I, HindIII
ACTB	TP90-1 GFP (pDY0333)	293	CCCG GCTTC CTTTG TCC	294	AACC ACAA CTAG AATG CAGT GA	/56-FAM/T G CTA TTG C/ZEN/ T TTA TTT GTG GGC CCG /3IABkFQ/	406	None

ACTB	TP90-1 rc GFP (pDY0334)	295	CCCG GCTTC CTTTG TCC	296	GAAC TCCAC GCCG TTCA	/56-FAM/ CC ATG AAG A/ZE N/T CGA GTG CCG CAT CA/3I ABkF Q/	407	None
ACTB	PhiBT1 GFP (pDY0367)	297	CCCG GCTTC CTTTG TCC	298	AACC ACAA CTAG AATG CAGT GA	/56-FAM/T G CTA TTG C/ZEN/ T TTA TTT GTG GGC CCG /3IABk FQ/	406	None
ACTB	PhiBT1 rc GFP (pDY0368)	299	CCCG GCTTC CTTTG TCC	300	GAAC TCCAC GCCG TTCA	/56-FAM/ CC ATG AAG A/ZE N/T CGA GTG CCG CAT CA/3I ABkF Q/	407	None
LMNB 1	GFP (pDY0186)	301	TCCTT ATCA CGGT CCCG CTCG	302	GAAC TCCAC GCCG TTCA	/56-FAM/ CC ATG AAG A/ZE N/T CGA GTG CCG CAT CA/3I	407	Eco91I, HindIII

						ABkFQ/		
NOLC1	GFP (pDY0186)	303	CGTCGACAACGGTAGTG	304	GAACTCCACGCCGTTCA	/56-FAM/CCATGAAGA/ZEN/TCGAGTGCCGCAT CA/3IABkFQ/	407	Eco91I, HindIII
SUPT16 H	GFP (pDY0186)	305	TCGCGTGATTCTCGGAAC	306	GAACTCCACGCCGTTCA	/56-FAM/CC ATGAAGA/ZEN/T CGAGTGCCGCAT CA/3IABkFQ/	407	Eco91I, HindIII
SRRM2	GFP (pDY0186)	307	GGGCGGTAAGTGGTTAGTTT	308	GAACTCCACGCCGTTCA	/56-FAM/CCATGAAGA/ZEN/TCGAGTGCCGCAT CA/3IABkFQ/	407	Eco91I, HindIII
DEPDC4	GFP (pDY0186)	309	AAGAGGCGGAGCCAGTA	310	GAACTCCACGCCGTTCA	/56-FAM/CCATGAAGA/ZEN/T	407	Eco91I, HindIII

						CGA GTG CCG CAT CA/3I ABkF Q/		
NES	GFP (pDY01 86)	311	CTCCC TTCTC CCGG TGCCC	312	GAAC TCCAC GCCG TTCA	/56-FAM/C C GGC TTG T/ZEN/ C GAC GAC GGC G/3IAB kFQ/	405	Eco91I, HindIII
ACTB	ACTB HITI templat e GFP (pDY02 19)	313	CCCG GCTTC CTTTG TCC	314	GAAC TCCAC GCCG TTCA	/56-FAM/ CC ATG AAG A/ZE N/T CGA GTG CCG CAT CA/3I ABkF Q/	407	Eco91I
SRRM2	SRRM2 HITI templat e GFP (aRY01 82_A2)	315	GGGC GGTA AGTG GTTA GTTT	316	GAAC TCCAC GCCG TTCA	/56-FAM/ CC ATG AAG A/ZE N/T CGA GTG CCG CAT CA/3I ABkF Q/	407	Eco91I
NOLC1	NOLC1 HITI templat e GFP (aRY01 82_A3)	317	CGTC GACA ACGG TAGT G	318	GAAC TCCAC GCCG TTCA	/56-FAM/ CC ATG AAG A/ZE N/T	407	Eco91I

						CGA GTG CCG CAT CA/3I ABkF Q/		
DEPDC 4	DEPDC 4 HITI templat e GFP (aRY01 82_A5)	319	AAGA GGCG GAGC CAGT A	320	GAAC TCCAC GCCG TTCA	/56-FAM/ CC ATG AAG A/ZE N/T CGA GTG CCG CAT CA/3I ABkF Q/	407	Eco91I
NES	NES HITI templat e GFP (aRY01 82_A7)	321	CTCCC TTCTC CCGG TGCCC	322	GAAC TCCAC GCCG TTCA	/56-FAM/ CC ATG AAG A/ZE N/T CGA GTG CCG CAT CA/3I ABkF Q/	407	Eco91I
LMNB 1	LMNB 1 HITI templat e GFP (aRY01 82_A4)	323	TCCTT ATCA CGGT CCCG CTCG	324	GAAC TCCAC GCCG TTCA	/56-FAM/ CC ATG AAG A/ZE N/T CGA GTG CCG CAT CA/3I ABkF Q/	407	Eco91I

ACTB	SERPINA (pDY0298)	325	CCCG GCTTC CTTTG TCC	326	GGCC TGCC AGCA GGAG GA	/56-FAM/ CC GGC TTG T/ZE N/C GAC GAC GGC G/3I ABkF Q/	405	EcoRI, XhoI, HindIII
ACTB	CPS1 (pDY299)	327	CCCG GCTTC CTTTG TCC	328	GGTG TGCA GTCA CATTG GTAA AGCC	/56-FAM/ AC AGC TTT C/ZE N/A AAG TGG TGA GGA CAC T/3IA BkFQ /	408	XhoI, HindIII
ACTB	CFTR (pDY0373)	329	CCCG GCTTC CTTTG TCC	330	GATG GGTCT AGTC CAGC TAAA G	/56-FAM/ TAC GGT ACA/ ZEN/ AAC CC ACC CGA GAG A/3I ABkF Q/	409	Eco91I, HindIII
ACTB	NYESO TRAC (pDY0318)	331	CCCG GCTTC CTTTG TCC	332	GAGA GACA AGGC TGCA CA	/56-FAM/ TAC GGT ACA/ ZEN/ AAC CC ACC CGA GAG A/3I	409	Eco47III, HindIII

						ABkFQ/		
NC_000003	GFP (pDY0186)	333	CCAGGTGAGAGTCAGGGTAGTGTTCA	334	GAACTCCACGCCGTTCA	/56-FAM/CCGGCTTGT/ZEN/CGACGACGGCG/3IABkFQ/	405	Eco91I, HindIII
NC_000002	GFP (pDY0186)	335	AGGGACCTTTGCCTGTGTGAGTC	336	GAACTCCACGCCGTTCA	/56-FAM/CCGGCTTGT/ZEN/CGACGACGGCG/3IABkFQ/	405	Eco91I, HindIII
NC_000009	GFP (pDY0186)	337	TCAGCTCTGTGCTGAGGCGAA	338	GAACTCCACGCCGTTCA	/56-FAM/CCGGCTTGT/ZEN/CGACGACGGCG/3IABkFQ/	405	Eco91I, HindIII
chr6:1490 45959	GFP (pDY0186)	339	AAGCCATCTCCCAGAATATCTGCTTAG	340	GAACTCCACGCCGTTCA	/56-FAM/CCGGCTTGT/ZEN/C	405	Eco91I, HindIII

			AAATG			GACGACGGCG/3IABkFQ/		
chr16:186 07730	GFP (pDY0186)	341	GAGAGGAGCAACAGTGAGCATGATG	342	GAACTCCACGCCGTTCA	/56-FAM/CCGGCTTGT/ZEN/CGACGACGGCG/3IABkFQ/	405	Eco91I, HindIII
chr6:1490 45959	ACTB HITI template GFP (pDY0219)	343	AAGCCATCTCCCAGAATATCTGCTTAGAAATG	344	GAACTCCACGCCGTTCA	/56-FAM/CCGGCTTGT/ZEN/CGACGACGGCG/3IABkFQ/	405	Eco91I
chr16:186 07730	ACTB HITI template GFP (pDY0219)	345	GAGAGGAGCAACAGTGAGCATGATG	346	GAACTCCACGCCGTTCA	/56-FAM/CCGGCTTGT/ZEN/CGACGACGGCG/3IABkFQ/	405	Eco91I
ACTB	CAG_Kozak_bGH_therapeutic_genes generic	347	CCCGGCTTCCTTTGTCC	348	GGCTATGAACTAATGACCCCGT	/56-FAM/CCGGCTTGT/ZEN/C	405	Eco91I, HindIII

	minicircle					GACGACGGCG/3IABkFQ/		
ACTB	Hibit-SERPINA (pDY0405)	349	CCCGGCTTCCTTTGTCC	350	GGCCTGCCAGCAGGAGGA	/56-FAM/CCGGCTTGT/ZEN/CGACGACGGCG/3IABkFQ/	405	EcoRI, XhoI, HindIII
ACTB	Hibit-CPS1 (pDY406)	351	CCCGGCTTCCTTTGTCC	352	GGTGTGCAGTCACATTGGTAAAGCC	/56-FAM/ACAGCTTTC/ZEN/AAAGTGGTGAGGACAC T/3IABkFQ/	408	XhoI, HindIII

[313] Sequences of primers used for NGS readout can be found in Table 8 below.

Table 8

SEQ ID NO / DESCRIPTION / SOURCE	**ID**	**SEQUENCE (5'-3')**
SEQ ID NO: 353 N-term ACTB Tn5 readout F 1 (Artificial Sequence)	PD0966	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCCGACCTCGGC TCACAGCG

SEQ ID NO: 354 N-term ACTB Tn5 readout F 2 (Artificial Sequence)	PD0967	ACACTCTTTCCCTACACGACGCTCTTCCGATCTACCGACCTCGG CTCACAGCG
SEQ ID NO: 355 N-term ACTB Tn5 readout F 3 (Artificial Sequence)	PD0968	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGACCGACCTCG GCTCACAGCG
SEQ ID NO: 356 N-term ACTB Tn5 readout F 4 (Artificial Sequence)	PD0969	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTGACCGACCTC GGCTCACAGCG
SEQ ID NO: 357 N-term ACTB Tn5 readout F 5 (Artificial Sequence)	PD0970	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCTGACCGACCT CGGCTCACAGCG
SEQ ID NO: 358 N-term ACTB Tn5 readout F 6 (Artificial Sequence)	PD0971	ACACTCTTTCCCTACACGACGCTCTTCCGATCTACTGACCGACC TCGGCTCACAGCG
SEQ ID NO: 359 N-term ACTB Tn5 readout F 7 (Artificial Sequence)	PD0972	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTACTGACCGAC CTCGGCTCACAGCG
SEQ ID NO: 360 N-term ACTB Tn5 readout F 8 (Artificial Sequence)	PD0973	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGTACTGACCGA CCTCGGCTCACAGCG

SEQ ID NO: 361 ACTB N-term NGS R for Cas14 indels (Artificial Sequence)	FP0952	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCCAC CCAGCC AGCTCCC
SEQ ID NO: 362 NGS EMX1 Forward 1 (Artificial Sequence)	PD0313	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCCGGT GGCGCAT TGCCAC
SEQ ID NO: 363 NGS EMX1 Forward 2 (Artificial Sequence)	PD0314	ACACTCTTTCCCTACACGACGCTCTTCCGATCTACCGG TGGCGCA TTGCCAC
SEQ ID NO: 364 NGS EMX1 Forward 3 (Artificial Sequence)	PD0315	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGACCG GTGGCGC ATTGCCAC
SEQ ID NO: 365 NGS EMX1 Forward 4 (Artificial Sequence)	PD0316	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTGACC GGTGGCG CATTGCCAC
SEQ ID NO: 366 NGS EMX1 Forward 5 (Artificial Sequence)	PD0317	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCTGAC CGGTGGC GCATTGCCAC
SEQ ID NO: 367 NGS EMX1 Forward 6 (Artificial Sequence)	PD0318	ACACTCTTTCCCTACACGACGCTCTTCCGATCTACTGA CCGGTGG CGCATTGCCAC

SEQ ID NO: 368 NGS EMX1 Forward 7 (Artificial Sequence)	PD0319	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTACTG ACCGGTG GCGCATTGCCAC
SEQ ID NO: 369 NGS EMX1 Forward 8 (Artificial Sequence)	PD0320	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGTACT GACCGGT GGCGCATTGCCAC
SEQ ID NO: 370 NGS EMX1 Reverse (Artificial Sequence)	PD0321	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCAGA GTCCAGC TTGGGCCCA

[314] Sequences of off-target sites can be found in Table 9 below.

Table 9

SEQ ID NO / DESCRIPTION / SOURCE	**SEQUENCE (5'-3')**
SEQ ID NO: 371 Cas9_chr6:149045959 (Artificial Sequence)	GATATTTTCCCAGCTCACCA
SEQ ID NO: 372 Cas9_chr16:18607730 (Artificial Sequence)	TCTATTCTCCCAGCTCCCCA
SEQ ID NO: 373 Bxb1_NC_000002 (Artificial Sequence)	AGCGGCTTCTGTCTCTGTGAGTGAGCTGGCGGTCTCCGTC

SEQ ID NO: 374 Bxb1_NC_000003 (Artificial Sequence)	GACTAGCCCACGCTCCGGTTCTGAGCCGCGACGGCGGTCTCCG
SEQ ID NO: 375 Bxb1_NC_000009 (Artificial Sequence)	CCCAGGGTCCCATGCGCTCCCCGGCCCTGACGGCGGTCTCC

[315] Linker sequences in Table 10 below.

Table 10

Description	Sequence (5'-3')	Amino acid sequence
A - P2A	GGAAGCGGAGCTACTAACTTCAGCCT GCTGAAGCAGGCTGGCGACGTGGAGG AGAACCCTGGACCT (SEQ ID NO: 410)	GSGATNFSLLKQAGDVEENPGP (SEQ ID NO: 418)
B - (GGGS)3	GGGGGAGGAGGTTCTGGAGGCGGAGG CTCCGGAGGCGGAGGGTCA (SEQ ID NO: 411)	GGGGSGGGGSGGGGS (SEQ ID NO: 419)
C - GGGGS	GGAGGTGGCGGGAGC (SEQ ID NO: 412)	GGGGS (SEQ ID NO: 420)
D - PAPAP	CCCGCACCAGCGCCT (SEQ ID NO: 413)	PAPAP (SEQ ID NO: 421)
E - (EAAAK)3	GAGGCAGCTGCCAAGGAAGCCGCT GCCAAGGAGGCGGCCGCAAAG (SEQ ID NO: 414)	EAAAKEAAAKEAAAK (SEQ ID NO: 422)
F - XTEN	AGTGGGAGCGAGACCCCTGGGACT AGCGAGTCAGCTACACCCGAAAGC (SEQ ID NO: 415)	SGSETPGTSESATPES (SEQ ID NO: 423)
G - (GGS)6	GGGGGGTCAGGTGGATCCGGCGG AAGTGGCGGATCCGGTGGATCTGG CGGCAGT (SEQ ID NO: 416)	GGSGGSGGSGGSGGSGGS (SEQ ID NO 424)
H - EAAAK	GAAGCTGCTGCTAAG (SEQ ID NO: 417)	EAAAK (SEQ ID NO: 425)

[316] Exemplary fusion sequences in Table 11 below.

Description	Sequence
SpCas9-XTEN-RT(1-478)-Sto7d-GGGGS-BxbINT Amino acid SEQ ID NO: 376	MKRTADGSEFESPKKKRKV**DKKYSIGLDIGTNSVGWAVITDEYKVPS KKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRR KNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNI VDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHF LIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSAR LSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDA KLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVN TEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQS KNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRK QRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIP YYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIER MTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAF LSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIE**

	ERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGK TILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHI ANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTT QKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYL QNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDK NRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERG GLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIRE VKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALI KKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMN FFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMP QVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDS PTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLE AKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELAL PSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISE FSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPA AFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDS GGSSGGSSGSETPGTSESATPESSGSETPGTSESATPESSGSETPGTSESAT PESSGGSSGGSST**LNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAET GGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLD QGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTV PNPYNLLSGPPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDP EMGISGQLTWTRLPQGFKNSPTLFNEALHRDLADFRIQHPDLILLQ YVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQV KYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGKAGFC RLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPA LGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDP VAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVK QPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEG LQHNCLDGTGGGGVTVKFKYKGEELEVDISKIKKVWRVGKMISFT YDDNGKTGRGAVSEKDAPKELLQMLEKSGKKSGGSKRTADGS**EFE PKKKRKVGGGGSPKKKRKVYPYDVPDYAG**SRALVVIRLSRVTDATTS PERQLESCQQLCAQRGWDVVGVAEDLDVSGAVDPFDRKRRPNLAR WLAFEEQPFDVIVAYRVDRLTRSIRHLQQLVHWAEDHKKLVVSAT EAHFDTTTPFAAVVIALMGTVAQMELEAIKERNRSAAHFNIRAGKY RGSLPPWGYLPTRVDGEWRLVPDPVQRERILEVYHRVVDNHEPLH LVAHDLNRRGVLSPKDYFAQLQGREPQGREWSATALKRSMISEAM LGYATLNGKTVRDDDGAPLVRAEPILTREQLEALRAELVKTSRAKP AVSTPSLLLRVLFCAVCGEPAYKFAGGGRKHPRYRCRSMGFPKHC GNGTVAMAEWDAFCEEQVLDLLGDAERLEKVWVAGSDSAVELAE VNAELVDLTSLIGSPAYRAGSPQREALDARIAALAARQEELEGLEAR PSGWEWRETGQRFGDWWREQDTAAKNTWLRSMNVRLTFDVRGG LTRTIDFGDLQEYEQHLRLGSVVERLHTGMS**
SpCas9-XTEN-RT(1-478)-Sto7d-GGGGS-BxbINT Nucleic acid SEQ ID NO: 377	ATGAAACGGACAGCCGACGGAAGCGAGTTCGAGTCACCAAAGAAG AAGCGGAAAGTCGACAAGAAGTACAGCATCGGCCTGGACATCGGCA CCAACTCTGTGGGCTGGGCCGTGATCACCGACGAGTACAAGGTGCC CAGCAAGAAATTCAAGGTGCTGGGCAACACCGACCGGCACAGCATC AAGAAGAACCTGATCGGAGCCCTGCTGTTCGACAGCGGCGAAACAG CCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATACACCA GACGGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAACGA GATGGCCAAGGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCC TTCCTGGTGGAAGAGGATAAGAAGCACGAGCGGCACCCCATCTTCG GCAACATCGTGGACGAGGTGGCCTACCACGAGAAGTACCCCACCAT

	CTACCACCTGAGAAAGAAACTGGTGGACAGCACCGACAAGGCCGAC CTGCGGCTGATCTATCTGGCCCTGGCCCACATGATCAAGTTCCGGGG CCACTTCCTGATCGAGGGCGACCTGAACCCCGACAACAGCGACGTG GACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGA GGAAAACCCCATCAACGCCAGCGGCGTGGACGCCAAGGCCATCCTG TCTGCCAGACTGAGCAAGAGCAGACGGCTGGAAAATCTGATCGCCC AGCTGCCCGGCGAGAAGAAGAATGGCCTGTTCGGAAACCTGATTGC CCTGAGCCTGGGCCTGACCCCCAACTTCAAGAGCAACTTCGACCTGG CCGAGGATGCCAAACTGCAGCTGAGCAAGGACACCTACGACGACGA CCTGGACAACCTGCTGGCCCAGATCGGCGACCAGTACGCCGACCTG TTTCTGGCCGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACAT CCTGAGAGTGAACACCGAGATCACCAAGGCCCCCCTGAGCGCCTCT ATGATCAAGAGATACGACGAGCACCACCAGGACCTGACCCTGCTGA AAGCTCTCGTGCGGCAGCAGCTGCCTGAGAAGTACAAAGAGATTTT CTTCGACCAGAGCAAGAACGGCTACGCCGGCTACATTGACGGCGGA GCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCATCCTGGAAA AGATGGACGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAGG ACCTGCTGCGGAAGCAGCGGACCTTCGACAACGGCAGCATCCCCCA CCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAA GATTTTTACCCATTCCTGAAGGACAACCGGGAAAAGATCGAGAAGA TCCTGACCTTCCGCATCCCCTACTACGTGGGCCCTCTGGCCAGGGGA AACAGCAGATTCGCCTGGATGACCAGAAAGAGCGAGGAAACCATCA CCCCCTGGAACTTCGAGGAAGTGGTGGACAAGGGCGCTTCCGCCCA GAGCTTCATCGAGCGGATGACCAACTTCGATAAGAACCTGCCCAAC GAGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCACCG TGTATAACGAGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAG AAAGCCCGCCTTCCTGAGCGGCGAGCAGAAAAAGGCCATCGTGGAC CTGCTGTTCAAGACCAACCGGAAAGTGACCGTGAAGCAGCTGAAAG AGGACTACTTCAAGAAAATCGAGTGCTTCGACTCCGTGGAAATCTCC GGCGTGGAAGATCGGTTCAACGCCTCCCTGGGCACATACCACGATC TGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGGAAAA CGAGGACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGG ACAGAGAGATGATCGAGGAACGGCTGAAAACCTATGCCCACCTGTT CGACGACAAAGTGATGAAGCAGCTGAAGCGGCGGAGATACACCGG CTGGGGCAGGCTGAGCCGGAAGCTGATCAACGGCATCCGGGACAAG CAGTCCGGCAAGACAATCCTGGATTTCCTGAAGTCCGACGGCTTCGC CAACAGAAACTTCATGCAGCTGATCCACGACGACAGCCTGACCTTT AAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAGGGCGATAGCC TGCACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAA GGGCATCCTGCAGACAGTGAAGGTGGTGGACGAGCTCGTGAAAGTG ATGGGCCGGCACAAGCCCGAGAACATCGTGATCGAAATGGCCAGAG AGAACCAGACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGAA TGAAGCGGATCGAAGAGGGCATCAAAGAGCTGGGCAGCCAGATCCT GAAAGAACACCCCGTGGAAAACACCCAGCTGCAGAACGAGAAGCT GTACCTGTACTACCTGCAGAATGGGCGGGATATGTACGTGGACCAG GAACTGGACATCAACCGGCTGTCCGACTACGATGTGGACGCTATCG TGCCTCAGAGCTTTCTGAAGGACGACTCCATCGACAACAAGGTGCT GACCAGAAGCGACAAGAACCGGGGCAAGAGCGACAACGTGCCCTC CGAAGAGGTCGTGAAGAAGATGAAGAACTACTGGCGGCAGCTGCTG AACGCCAAGCTGATTACCCAGAGAAAGTTCGACAATCTGACCAAGG CCGAGAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAA GAGACAGCTGGTGGAAACCCGGCAGATCACAAAGCACGTGGCACA GATCCTGGACTCCCGGATGAACACTAAGTACGACGAGAATGACAAG

	CTGATCCGGGAAGTGAAAGTGATCACCCTGAAGTCCAAGCTGGTGT CCGATTTCCGGAAGGATTTCCAGTTTTACAAAGTGCGCGAGATCAAC AACTACCACCACGCCCACGACGCCTACCTGAACGCCGTCGTGGGAA CCGCCCTGATCAAAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTA CGGCGACTACAAGGTGTACGACGTGCGGAAGATGATCGCCAAGAGC GAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACAGCA ACATCATGAACTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGA GATCCGGAAGCGGCCTCTGATCGAGACAAACGGCGAAACCGGGGA GATCGTGTGGGATAAGGGCCGGGATTTTGCCACCGTGCGGAAAGTG CTGAGCATGCCCCAAGTGAATATCGTGAAAAAGACCGAGGTGCAGA CAGGCGGCTTCAGCAAAGAGTCTATCCTGCCCAAGAGGAACAGCGA TAAGCTGATCGCCAGAAAGAAGGACTGGGACCCTAAGAAGTACGGC GGCTTCGACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGGCCAA AGTGGAAAAGGGCAAGTCCAAGAAACTGAAGAGTGTGAAAGAGCT GCTGGGGATCACCATCATGGAAAGAAGCAGCTTCGAGAAGAATCCC ATCGACTTTCTGGAAGCCAAGGGCTACAAAGAAGTGAAAAAGGACC TGATCATCAAGCTGCCTAAGTACTCCCTGTTCGAGCTGGAAAACGGC CGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAGAAGGGAAACG AACTGGCCCTGCCCTCCAAATATGTGAACTTCCTGTACCTGGCCAGC CACTATGAGAAGCTGAAGGGCTCCCCCGAGGATAATGAGCAGAAAC AGCTGTTTGTGGAACAGCACAAGCACTACCTGGACGAGATCATCGA GCAGATCAGCGAGTTCTCCAAGAGAGTGATCCTGGCCGACGCTAAT CTGGACAAAGTGCTGTCCGCCTACAACAAGCACCGGGATAAGCCCA TCAGAGAGCAGGCCGAGAATATCATCCACCTGTTTACCCTGACCAAT CTGGGAGCCCCTGCCGCCTTCAAGTACTTTGACACCACCATCGACCG GAAGAGGTACACCAGCACCAAAGAGGTGCTGGACGCCACCCTGATC CACCAGAGCATCACCGGCCTGTACGAGACACGGATCGACCTGTCTC AGCTGGGAGGTGACTCTGGAGGATCTAGCGGAGGATCCTCTGGCAG CGAGACACCAGGAACAAGCGAGTCAGCAACACCAGAGAGCTCTGGT AGCGAGACACCCGGTACCAGTGAAAGCGCCACGCCAGAAAGCAGT GGGAGTGAGACTCCGGGTACATCTGAATCAGCGACACCGGAATCAA GTGGCGGCAGCAGCGGCGGCAGCAGCACCCTAAATATAGAAGATGA GTATCGGCTACATGAGACCTCAAAAGAGCCAGATGTTTCTCTAGGGT CCACATGGCTGTCTGATTTTCCTCAGGCCTGGGCGGAAACCGGGGGC ATGGGACTGGCAGTTCGCCAAGCTCCTCTGATCATACCTCTGAAAGC AACCTCTACCCCCGTGTCCATAAAACAATACCCCATGTCACAAGAA GCCAGACTGGGGATCAAGCCCCACATACAGAGACTGTTGGACCAGG GAATACTGGTACCCTGCCAGTCCCCCTGGAACACGCCCCTGCTACCC GTTAAGAAACCAGGGACTAATGATTATAGGCCTGTCCAGGATCTGA GAGAAGTCAACAAGCGGGTGGAAGACATCCACCCCACCGTGCCCAA CCCTTACAACCTCTTGAGCGGGCCCCCACCGTCCCACCAGTGGTACA CTGTGCTTGATTTAAAGGATGCCTTTTTCTGCCTGAGACTCCACCCC ACCAGTCAGCCTCTCTTCGCCTTTGAGTGGAGAGATCCAGAGATGGG AATCTCAGGACAATTGACCTGGACCAGACTCCCACAGGGTTTCAAA AACAGTCCCACCCTGTTTAATGAGGCACTGCACAGAGACCTAGCAG ACTTCCGGATCCAGCACCCAGACTTGATCCTGCTACAGTACGTGGAT GACTTACTGCTGGCCGCCACTTCTGAGCTAGACTGCCAACAAGGTAC TCGGGCCCTGTTACAAACCCTAGGGAACCTCGGGTATCGGGCCTCG GCCAAGAAAGCCCAAATTTGCCAGAAACAGGTCAAGTATCTGGGGT ATCTTCTAAAAGAGGGTCAGAGATGGCTGACTGAGGCCAGAAAAGA GACTGTGATGGGGCAGCCTACTCCGAAGACCCCTCGACAACTAAGG GAGTTCCTAGGGAAGGCAGGCTTCTGTCGCCTCTTCATCCCTGGGTT TGCAGAAATGGCAGCCCCCCTGTACCCTCTCACCAAACCGGGGACT

	CTGTTTAATTGGGGCCCAGACCAACAAAAGGCCTATCAAGAAATCA AGCAAGCTCTTCTAACTGCCCCAGCCCTGGGGTTGCCAGATTTGACT AAGCCCTTTGAACTCTTTGTCGACGAGAAGCAGGGCTACGCCAAAG GTGTCCTAACGCAAAAACTGGGACCTTGGCGTCGGCCGGTGGCCTA CCTGTCCAAAAAGCTAGACCCAGTAGCAGCTGGGTGGCCCCCTTGC CTACGGATGGTAGCAGCCATTGCCGTACTGACAAAGGATGCAGGCA AGCTAACCATGGGACAGCCACTAGTCATTCTGGCCCCCCATGCAGTA GAGGCACTAGTCAAACAACCCCCCGACCGCTGGCTTTCCAACGCCC GGATGACTCACTATCAGGCCTTGCTTTTGGACACGGACCGGGTCCAG TTCGGACCGGTGGTAGCCCTGAACCCGGCTACGCTGCTCCCACTGCC TGAGGAAGGGCTGCAACACAACTGCCTTGATGGGACAGGTGGCGGT GGTGTCACCGTCAAGTTCAAGTACAAGGGTGAGGAACTTGAAGTTG ATATTAGCAAAATCAAGAAGGTTTGGCGCGTTGGTAAAATGATATC TTTTACTTATGACGACAACGGCAAGACAGGTAGAGGGGCAGTGTCT GAGAAAGACGCCCCCAAGGAGCTGTTGCAAATGTTGGAAAAGTCTG GGAAAAAGTCTGGCGGCTCAAAAAGAACCGCCGACGGCAGCGAATT CGAGCCCAAGAAGAAGAGGAAAGTCGGAGGTGGCGGGAGCCCAAA AAAGAAAAGAAAAGTGTATCCCTATGATGTCCCCGATTATGCCGGT TCAAGAGCCCTGGTCGTGATTAGACTGAGCCGAGTGACAGACGCCA CCACAAGTCCCGAGAGACAGCTGGAATCATGCCAGCAGCTCTGTGC TCAGCGGGGTTGGGATGTGGTCGGCGTGGCAGAGGATCTGGACGTG AGCGGGGCCGTCGATCCATTCGACAGAAAGAGGAGGCCCAACCTGG CAAGATGGCTCGCTTTCGAGGAACAGCCCTTTGATGTGATCGTCGCC TACAGAGTGGACCGGCTGACCCGCTCAATTCGACATCTCCAGCAGCT GGTGCATTGGGCTGAGGACCACAAGAAACTGGTGGTCAGCGCAACA GAAGCCCACTTCGATACTACCACACCTTTTGCCGCTGTGGTCATCGC ACTGATGGGCACTGTGGCCCAGATGGAGCTCGAAGCTATCAAGGAG CGAAACAGGAGCGCAGCCCATTTCAATATTAGGGCCGGTAAATACA GAGGCTCCCTGCCCCCTTGGGGATATCTCCCTACCAGGGTGGATGGG GAGTGGAGACTGGTGCCAGACCCCGTCCAGAGAGAGCGGATTCTGG AAGTGTACCACAGAGTGGTCGATAACCACGAACCACTCCATCTGGT GGCACACGACCTGAATAGACGCGGCGTGCTCTCTCCAAAGGATTAT TTTGCTCAGCTGCAGGGAAGAGAGCCACAGGGAAGAGAATGGAGTG CTACTGCACTGAAGAGATCTATGATCAGTGAGGCTATGCTGGGTTAC GCAACACTCAATGGCAAAACTGTCCGGGACGATGACGGAGCCCCTC TGGTGAGGGCTGAGCCTATTCTCACCAGAGAGCAGCTCGAAGCTCT GCGGGCAGAACTGGTCAAGACTAGTCGCGCCAAACCTGCCGTGAGC ACCCCAAGCCTGCTCCTGAGGGTGCTGTTCTGCGCCGTCTGTGGAGA GCCAGCATACAAGTTTGCCGGCGGAGGGCGCAAACATCCCCGCTAT CGATGCAGGAGCATGGGGTTCCCTAAGCACTGTGGAAACGGGACAG TGGCCATGGCTGAGTGGGACGCCTTTTGCGAGGAACAGGTGCTGGA TCTCCTGGGTGACGCTGAGCGGCTGGAAAAAGTGTGGGTGGCAGGA TCTGACTCCGCTGTGGAGCTGGCAGAAGTCAATGCCGAGCTCGTGG ATCTGACTTCCCTCATCGGATCTCCTGCATATAGAGCTGGGTCCCCA CAGAGAGAAGCTCTGGACGCACGAATTGCTGCACTCGCTGCTAGAC AGGAGGAACTGGAGGGCCTGGAGGCCAGGCCCTCTGGATGGGAGTG GCGAGAAACCGGACAGAGGTTTGGGGATTGGTGGAGGGAGCAGGA CACCGCAGCCAAGAACACATGGCTGAGATCCATGAATGTCCGGCTC ACATTCGACGTGCGCGGTGGCCTGACTCGAACCATCGATTTTGGCGA CCTGCAGGAGTATGAACAGCACCTGAGACTGGGGTCCGTGGTCGAA AGACTGCACACTGGGATGTCC

SpCas9 Amino acid SEQ ID NO: 378	DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFH RLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDK ADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFE ENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLG LTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKN LSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPE KYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLN REDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILT FRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIER MTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSG EQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASL GTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAH LFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGIL QTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIE EGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRL SDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKN YWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITK HVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREI NNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKS EQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWD KGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKK DWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERS SFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQ KGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEII EQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGA PAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD
RT(1-478)-Sto7d Amino acid SEQ ID NO: 379	LNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLII PLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLP VKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGPPPSHQWYTV LDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSPT LFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQT LGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPT PKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQK AYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRR PVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAP HAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLL PLPEEGLQHNCLDGTGGGGVTVKFKYKGEELEVDISKIKKVWRVGKMI SFTYDDNGKTGRGAVSEKDAPKELLQMLEKSGKKSGGSKRTADGS
BxbINT Amino acid SEQ ID NO: 380	SRALVVIRLSRVTDATTSPERQLESCQQLCAQRGWDVVGVAEDLDVSG AVDPFDRKRRPNLARWLAFEEQPFDVIVAYRVDRLTRSIRHLQQLVHW AEDHKKLVVSATEAHFDTTTPFAAVVIALMGTVAQMELEAIKERNRSA AHFNIRAGKYRGSLPPWGYLPTRVDGEWRLVPDPVQRERILEVYHRVV DNHEPLHLVAHDLNRRGVLSPKDYFAQLQGREPQGREWSATALKRSM ISEAMLGYATLNGKTVRDDDGAPLVRAEPILTREQLEALRAELVKTSRA KPAVSTPSLLLRVLFCAVCGEPAYKFAGGGRKHPRYRCRSMGFPKHCG NGTVAMAEWDAFCEEQVLDLLGDAERLEKVWVAGSDSAVELAEVNA ELVDLTSLIGSPAYRAGSPQREALDARIAALAARQEELEGLEARPSGWE WRETGQRFGDWWREQDTAAKNTWLRSMNVRLTFDVRGGLTRTIDFG DLQEYEQHLRLGSVVERLHTGMS

EXAMPLES

[317] While several experimental Examples are contemplated, these Examples are intended to be non-limiting.

Example 1

CRE Integration Efficiency

[318] The efficiency of the CRE integration was tested. In order to test the efficacy of PASTE with GFP using lox71/lox66/Cre recombinase system, a clonal HEK293FT cell line with lox71 sequence (SEQ ID NO: 1) integrated into the genome using lentivirus was developed. The integration of GFP was tested by transfection of modified HEK293FT cell line with: (1) plus/minus SEQ ID NO: 71 comprising a Cre recombinase expression plasmid, and (2) SEQ ID NO: 72 comprising a GFP template and a lox 66 Cre site of SEQ ID NO: 2. After 72 hours, the percent integration of GFP into the lox71 site was probed. FIG. 3 shows the percent integration of GFP in the lentiviral integrated lox71 site in HEK293FT cell line in the presence of various plasmids. It was observed that pCMV PE2 P2A Cre (SEQ ID NO: 73), a mammalian expression vector with prime editing complex and Cre recombinase linked to PE2 via a cleavable linker or a non-cleavable linker, shows integration of GFP.

Example 2

Programmable Addition via Site-Specific Targeting Elements (PASTE) with Cre Recombinase - Addition of Lox Site

[319] The lox71 (SEQ ID NO: 1) or lox66 (SEQ ID NO: 2) sequence was inserted into the HEK293FT cell genome using prime editing to test integration of GFP into the HEK293FT genome. In order to insert lox71 or lox66 sequence into HEK293FT cell genome, a pegRNA with PBS length of 13 base pairs operably linked to RT region of varying lengths was used. The following plasmids were used in the transfection of HEK293FT cells. The cells were transfected with (1) prime editing construct (PE2) or PE2 with conditional Cre expression, (2) Lox71 or Lox66 pegRNA targeting the HEK3 locus, and (3) plus/minus +90 HEK3 nicking second guide RNA targeting the HEK3 locus (+90 ngRNA). After 72 hours, the percent editing of the HEK293FT genome at the HEK3 locus was probed for incorporation of various lengths of lox71 or lox66 (see FIG. 4). It was observed that 34 base pair lox71 (HEK3 locus guide, SEQ ID NO: 83; and Lox71 pegRNA with RT 34 and PBS 13, SEQ ID NO: 81) with

+90 ngRNA (SEQ ID NO: 75) and 34 base pair lox66 (HEK3 locus guide, SEQ ID NO: 83; and Lox66 pegRNA with RT 34 and PBS 13, SEQ ID NO: 82) with +90 ngRNA (SEQ ID NO: 75) had the highest percent editing.

Example 3

PASTE with Cre Recombinase - Integration of Gene

[320] The lox71 or lox66 pegRNAs having PBS length of 13 base pairs and insert length of 34 base pairs were used to probe integration of GFP in the HEK293F genome. The PE and Cre were delivered in an inducible expression vectors and induced at day 2. The HEK293FT cells were transfected with the following plasmids: (1) prime editing construct (PE2 or PE2 with conditional Cre expression); (2) Lox71 pegRNA; (3) plus/minus +90 HEK3 nicking guide RNA; and (4) EGFP template with Lox66 site. After 72 hours, the percent editing of lox71 site and percent integration of GFP was probed with or without lox66 site in the presence of various PE/Cre constructs. FIG. 5A summarizes the percent editing of lox71 site with different PE/Cre vectors. FIG. 5B summarizes the percent integration of GFP at the lox71 site in HEK293FT cell genome. It was observed that although the lox71 site was edited in the presence of inducible or non-inducible PE/Cre expression system, there was no GFP integration.

Example 4

Bxb1 Integration Data Lenti Reporter

[321] The integration system was switched to an integrase system that could result in an integration of target genes into a genome with higher efficiency. Serine integrase Bxb1 has been shown to be more active than Cre recombinase and highly efficient in bacteria and mammalian cells for irreversible integration of target genes. FIG. 6 shows a schematic of PASTE methodology using Bxb1 (Merrick, C. A. *et al., ACS Synth. Biol.* 2018, 7, 299−310).

[322] To probe the efficiency of the Bxb1 integration system, a clonal HEK293FT cell line with attB Bxb1 site (SEQ ID NO: 3) integrated using lentivirus was developed. The modified HEK293FT cell line was then transferred with the following plasmids: (1) plus/minus Bxb1 expression plasmid and (2) plus/minus GFP (SEQ ID NO: 76) or G-Luc (SEQ ID NO: 77) minicircle template with attP Bxb1 site. After 72 hours, the integration of GFP or Gluc into the attB site in the HEK293FT genome was probed. The percent integrations of GFP or Gluc

into the attB locus are shown in FIG. 7. It was observed that GFP and Gluc showed efficient integration into the attB site in HEK293FT cells.

Example 5

Addition of Bxb1 Site to Human Genome Using PRIME

[323] The maximum length of attB that can be integrated into a HEK293FT cell line with the best efficiency was probed. To probe the best length of attB (SEQ ID NO: 3) or its reverse complement attP (SEQ ID NO: 4) for prime editing, pegRNAs having PBS length of 13 nt with varying RT homology length were used. The following plasmids were transfected in HEK293FT: (1) prime expression plasmid; (2) HEK3 targeting pegRNA design; and (3) HEK3 +90 nicking guide. After 72 hours, the percent integration of each of the attB construct was probed. FIG. 8 shows the percent editing in each HEK3 targeting pegRNA. It was observed that attB with 44, 34 and 26 base pairs and attB reverse complement with 34 and 26 base pairs showed the highest percent editing.

[324] Integration PASTE was then tested with tagging cell-organelle marker proteins with GFP in HEK29FT cells. PASTE was used to tag SUPT16H, SRRM2, LAMNB1, NOLC1 and DEPDC4 with GFP in different cell-culture wells and to test the usefulness of PASTE in tracking protein localization within the cells using microscopy. FIGS. 9A-9G shows the fluorescent microscopy results for each of the organelles. SUPT16H-GFP was observed to be enriched in the nucleus, SRRM2-GFP was observed to be enriched in the nuclear speckles, LAMNB1-GFP was observed to be enriched in the nuclear membrane, NOLC1-GFP was observed to be enriched in the fibrillar center, and DEPDC4-GFP was observed to be enriched in the aggresome.

[325] The transfection of the plasmids can be achieved using electroporation as illustrated in FIGS. 10A-10B.

Example 6

Programmable Integration of Genes with PASTE

[326] The efficiency of gene integration of Gluc or EGFP with PASTE was tested. To enable gene integration with PASTE, the following HEK3 targeting pegRNAs were used: (1) 44 pegRNA: PBS of 13nt and RT homology of 44nt; (2) 34 pegRNA: PBS of 13nt and RT homology of 34nt; and (3) 26 pegRNA: PBS of 13nt and RT homology of 26nt.

[327] A HEK293 cell line was transfected with following plasmids HEK293FT: (1) Prime expression plasmid; (2) Bxb1 expression plasmid; (3) HEK3 targeting pegRNA design; (4) HEK3 +90 nicking guide; and (5) EGFP or Gluc minicircle. After 72 hours, the percent integration of Gluc or EGFP was observed. FIG. 11 shows integration of EGFP and Gluc with each of the tested HEK3 targeting pegRNAs. It was observed that EGFP and Gluc were efficiently integrated using PASTE.

Example 7

PASTE for Integration of Multiple Genes

[328] The PASTE technique for site-specific integration of multiple genes into a cell is facilitated with the use of orthogonal attB and attP sites. Central dinucleotide can be changed to GA from GT, and only GA containing attB/attP sites can interact and do not cross react with GT containing sequences. A screen of dinucleotide combinations to find orthogonal attB/attP pairs for multiplexed PASTE editing can be performed. It has been shown that many orthogonal dinucleotide combinations can be found using a Bxb1 reporter system.

[329] To test this, $attB^{GT}$ and $attB^{GA}$ dinucleotides for Bxb1 was added at a ACTB site by prime editing. A EGFP - $attP^{GT}$ DNA minicircle and a mCherry - $attP^{GA}$ DNA minicircle was introduced to test the percent EGFP and mCherry editing in the presence or absence of Bxb1. The results of EGFP and mCherry editing are shown in FIGS. 14A-14B.

[330] Orthogonal editing with the right GT-EGFP and GA-mCherry pairs was achieved demonstrating the ability for multiplexed PASTE editing in cells.

[331] Two genes were introduced in the same cell using multiplexed PASTE to tag two different genes in a single reaction. EGFP and mCherry were tagged into the loci of ACTB and NOLC1 in a x cell line, in a single reaction. Further, EGFP and mCherry were tagged into the loci of ACTB and LAMNB1. The cells were visualized using fluorescence microscopy. FIGS. 15A-15B show the results of fluorescent microscopy for multiplexed PASTE.

[332] The ability of multiplexing with 9-different attB and attP central dinucleotides – AA, GA, CA, AG, AC, CC, GT, CT and TT (SEQ ID NOs: 7, 8, 23, 24, 19, 20, 25, 26, 27, 28, 9, 10, 15, 16, 17, 18, 5 and 6) – in a 9x9 cross of attB and attP was tested. The edits were probed using next-generation sequencing. The results of the 9X9 cross of attB and attP central dinucleotides – AA, GA, CA, AG, AC, CC, GT, CT and TT – are shown in FIG. 16A. Only

orthogonal pairs of attB and attP show the highest edit percentage. This result is also shown in the heat-map of FIG. 16B.

Example 8

Integration of Albumin and CPS1 Into Albumin Locus

[333] 12 pegRNAs with albumin guide were linked to PBS and reverse transcriptase sequence of variable length, and different nicking guide RNAs were used to transfect HEK293FT cells. The percent editing in the albumin was probed using next-generation sequencing. The results of prime editing at the albumin locus are shown in FIG. 17. It was observed that SEQ ID NO: 79 showed the highest percent edits with SERPINA1 and SEQ ID NO: 80 showed the highest percent edits with CPS1.

Example 9

Engineering T-cells

[334] In order to engineer CD8+ T-cells, the efficiency of PASTE delivery and editing in T-cells can be evaluated (FIG. 18). ACTB targeting pegRNA can be used to insert an integration site with an EGFP insertion template. To deliver the PASTE components to CD8+ T-cells, electroporation can be used along with an optimized electroporation protocol for unstimulated T-cells. As multiple plasmids may reduce the efficiency of electroporation, the consolidated PASTE components that use fewer vectors can be applied.

[335] Five vectors, three vectors, and two vectors PASTE systems show that robust T-cell editing can be achieved with maximal editing using the three- vector approach (FIG. 19). Further, expanded sets of electroporation conditions, including the overall plasmid amounts, cell numbers, and voltage/amperage protocol can be tested. In addition, stimulation of T-cells may influence the efficiency of transduction and PASTE efficiency. Further, CD4+/CD8+ T cell mixtures stimulated with T-Activator CD3/CD28 ligands can have higher PASTE editing efficiency versus unstimulated cells. In order to separate efficiency of PASTE from the overall delivery rate, an mCherry expression cassette on PASTE vectors can be evaluated in order to sort successfully transfected T cells. Once optimized parameters are achieved, a panel of 10 insertion sites with PASTE in T cells, including the TRAC, IL2Rα, and PDCD1 loci, can be evaluated, using different insertions (e.g. EGFP, BFP, and YFP), both in single and multiplexed editing contexts. A tested subset of relevant sites in HEK293FT achieved greater than 40% editing for EGFP insertion (FIG. 20). The PASTE efficiency at TRAC locus

with different TCR and CAR constructs can be evaluated. The T-cells can successfully be transfected to achieve insertion of CARs or TCRs.

Example 10

PASTE for CFTR

[336] PASTE for the CFTR locus can be tested in HEK293FT cells to identify top performing pegRNA and nicking designs for human cells. Neuro-2A cells can also be tested to identify top performing pegRNA and nicking designs for mouse cells. The best constructs can be applied for testing in mouse air lung interface (ALI) organoids *in vitro* or for delivery in pre-clinical models of cystic fibrosis in mice. Table 12 shows the pegRNA, nicking guide and minicircle DNA characteristics for the CFTR gene modulation.

Table 12

Variables	**Characteristics**
pegRNA	38 bp shortened minimal attB and normal 46 bp attB sequence with: a. PBS of 17, 13, and 9 nt length, and b. RT of 20, 15, and 10 nt in length
Nicking guides	Nicking guide 1 +64 bp Nicking guide 2 +23bp Nicking guide 3 -60 bp Nicking guide 4 -78 bp (distance is calculated from cut site of pegRNA)
Minicircle template	A. CFTR coding sequence alone (~4,454 pb in size) B. CFTR coding sequence plus 5' and 3' UTRs (~6,011 bp in size) (Both minicircles have attP site on them for integration by Bxb1 and a bGH poly A signal)

Example 11

AttB and EGPF Integration Using PASTE

[337] The efficiency of the integration of attB and EGPF at the ACTB locus was evaluated (FIGS. 21A-21C). To investigate whether Bxb1 can add an EGFP template into this site, a delivery approach using a 5 plasmid system expressing each of the following component was deployed: 1) pegRNA expression, 2) nicking guide expression, 3) Prime expression (Cas9-RT), 4) Bxb1 expression and 5) the insertion template (in this case EGFP). This approach

was found to yield editing efficiency of the attB site up to 24% and integration of EGFP ~10% in HEK293FT cells as measured by sequencing (FIGS. 21A-21B). Optimal activity is achieved in 3- 4 days and can be performed as a single step transfection or electroporation of all components. Because the EGFP plasmid is designed as a minicircle, allowing removal of all undesired bacterial components, only the desired gene is inserted along with minimal scars from the Bxb1 recombined sites.

[338] To make the tool simpler to use, the Bxb1 can be linked to Prime via a P2A linker to the Cas9-RT fusion, allowing for only a single plasmid to be used for PASTE protein expression rather than two. This optimization can maintain the same level of editing, making it easier to use the tool and deliver it (FIG. 21C).

Example 12

Programmable EGFP Integrations in Different Cell Types

[339] The programmable EGFP integration in liver hepatocellular carcinoma cell line HEPG2 (FIG. 22A) and chronic myelogenous leukemia cell line K562 (FIG. 22B) was evaluated. EGFP integration at the ACTB locus in K562 and HEPG2 cells of about 15% was observed, demonstrating robustness of the platform across cell types.

Example 13

Mutagenesis of Bxb1 for Enhanced PASTE Activity

[340] The mutagenesis of Bxb1 for enhanced PASTE activity was evaluated (FIGS. 23A-23C). Two levers for optimizing PASTE activity exist: 1) improving the activity of the integrase and 2) enhancing the Prime addition of the integration sequence. As illustrated in FIGS. 23A-23B, Bxb1 activity can be improved as only about 30% of Bxb1 attB sites that are added by PASTE are integrated into by Bxb1. This illustrates that if the Bxb1 efficiency can be improved, the PASTE can be improved. Furthermore, catalytic residues in the Bxb1 integrase were identified via conservation and structural analyses and Bxb1 mutants were generated to test as part of PASTE. As illustrated in FIG. 23B, the mutations can improve integration by about 20-30%.

Example 14

Effect of the pegRNA PBS and RT Lengths on the Prime Editing Integration Efficiency

[341] The effect of the pegRNA PBS and RT lengths on the prime editing integration efficiency was evaluated (FIGS. 25A-25F). It was found that PASTE can be optimized by tuning the PBS and RT lengths at the ACTB locus to achieve editing rates up to about 20% (FIG. 25A). It was found that shortening the attB site can help improve PASTE function as Prime is better at inserting shorter sequences. Further optimization of PBS, RT, and attB lengths showed that optimal designs can be found for insertion upstream of the LMNB1, NOLC1, and GRSF1 loci (FIGS. 25B, 25C, and 25D). Lengths as short as 36nt for attB were found to be still functional for integration into a reporter plasmid (FIGS. 25B and 25C). It was found that the reverse complemented version of the attB sequence was better integrated via Prime editing, suggesting that the sequence of what Prime is inserting matters. EGFP integrations with attP site mutants showed that certain mutants can improve integration efficiency significantly (FIG. 25E). PASTE was also performed with a large panel of genes, inserting EGFP at the N-terminus of ACTB, LMNB1, SUPT16H, SRRM2, NOLC1, KLHL15, GRSF1, DEPDC4, NES, PGM1, CLTA, BASP1, and DNAJC18 (FIG. 25F). Editing rates that are about 5%-40% were found using digital droplet PCR (ddPCR).

Example 15

Comparison of PASTE and HITI On-target and Off-target Activities

[342] The PASTE and HITI on-target and off-target activities were compared (FIGS. 26A-26F). PASTE and HITI were found to have about 22% and 5% integration efficiencies respectively when using the same guide sequence (FIGS. 26A and 26B). PASTE was found to outperform HITI at most sites when analyzing the editing of 14 genes (FIG. 26C). Using a ddPCR based approach, it was found that PASTE was very specific with minimal off-target activity for Bxb1 off-targets integrations (FIG. 26D) and Cas9 off-targets integrations (FIG. 26E). The analysis of inserts of different sizes showed that PASTE can reliably insert sequences 1kb-10 kb in size (FIG. 26F), revealing the wide range of sequence sizes PASTE is capable of working with. A decrease in insertion efficiency at larger sizes was also observed, which was likely due to the reduction in plasmid delivery to HEK293FT cells at larger plasmid sizes.

Example 16

Multiplexing with PASTE and Orthogonal Di-nucleotide attB and attP Sites

[343] Multiplexing with PASTE and orthogonal di-nucleotide attB and attP sites was evaluated (FIGS. 28A-28C). Multiple orthogonal combinations were found for mutants of the central di-nucleotide motif (FIGS. 28A and 28B). As illustrated in FIG. 28C, programmable multiplexed gene insertion can be achieved by using these orthogonal combinations with PASTE only delivering different pegRNAs and gene inserts while keeping the protein components the same (Fig. 8C).

Example 17

PASTE Multiplexed Integrations at Endogenous Sites

[344] PASTE multiplexed integrations at endogenous sites were evaluated (FIGS. 28A-28G). A reading frame for the attR scar that is left post-integration by Bxb1 that is ideal for a protein linker due to the enrichment of glycines, serines, and prolines in the sequence (GLSGQPPRSPSSGSSG (SEQ ID NO: 426)) was identified. PegRNAs were designed using this linker frame for the resolution of the attR for tagging a number of genes at the N-terminus with EGFP (ACTB, NOLC1, LMNB1, SUPT16H, SRRM2, and DEPDC4). As these genes all have distinct protein localization appearances, microscopy can be used for ascertaining proper gene tagging. PASTE was found to be capable of high-efficiency gene tagging with protein localizations that match the reference images and expected localization of the proteins in the cells (FIGS. 28A-28C). Genes were also tagged in multiplexed fashion to demonstrate the orthogonality of the engineered integration sites. ACTB, LMNB1, NOLC1, and GRSF were targeted with orthogonal pegRNAs carrying GT, TG, AC, and CA, respectively in HEK293FT in groups of single, dual-plexing, and triple-plexing (FIGS. 28D-28E). These dinucleotides were paired with templates carrying EGFP, BFP, and mCherry to allow for multicolor imaging of these labeled genes. The efficiencies of integration for these multiplexing experiments were found to range from about 5%-32%, revealing efficient multiplex integration with PASTE. Using confocal microscopy of these multiplexed integration experiments, cells were found with simultaneous labeling of these different proteins (FIGS. 28F-28G).

Example 18

Combination of CRISPR-based Genome Editing and Site-specific Integration

[345] The combination of CRISPR-based genome editing and site-specific integration was evaluated.

[346] PegRNAs containing different attB length truncations were assessed (FIG. 29A). Prime editing was found to be capable of inserting sequences up to 56 bp at the beta-actin (*ACTB*) gene locus, with higher efficiency at lengths below 31 bp (FIGS. 29A-B) The integration of cognate landing sites was tested for multiple insertion enzymes: Bxb1, TP901, and phiBT1 phage serine integrases and Cre recombinase. Prime editing successfully inserted all landing sites tested, with efficiencies between 10-30% (FIGS. 29C-D). To test the complete system, all components were combined and delivered in a single transfection: the prime editing vector, the landing site containing pegRNA, a nicking guide for stimulating prime editing, a mammalian expression vector for the corresponding integrase or recombinase and a 969 bp minicircle DNA cargo encoding green fluorescent protein (*GFP*) (FIG. 29E). *GFP* integration rates among the four integrases and recombinases were compared and Bxb1 integrase was found to have the highest integration rate (~20%) at the targeted *ACTB* locus and require the prime editing nicking guide for optimal performance (FIGS. 29F-H). Finally, to reduce the number of transfected components, Bxb1 was co-expressed with the SpCas9-M-MLV reverse transcriptase (PE2) fusion protein via a P2A protein cleavage site. This combination maintained high *GFP* insertion efficiency, up to 30% (FIG. 29E). The complete system, PASTE, achieved precise integration of templates as large as 9,500 bp with greater than 10% integration efficiency (FIGS. 29J-K and 26E), with complete integration of the full-length cargo confirmed by Sanger sequencing (FIG. 30A-E).

Example 19

Impact of Prime Editing and Integrase Parameters on PRIME Editing

[347] The impact of prime editing and integrase parameters on the integration efficiency of PRIME editing was assessed.

[348] Relevant pegRNA parameters for PASTE include the primer binding site (PBS), reverse transcription template (RT), and attB site lengths, as well as the relative locations and efficacy of the pegRNA spacer and nicking guide (FIG. 31A). A range of PBS and RT lengths were tested at two loci, *ACTB* and lamin B1 (*LMNB1*), and rules governing efficiency were found to vary between loci, with shorter PBS lengths and longer RT designs having

higher editing at the *ACTB* locus (FIG. 31B) and longer PBS and shorter RT designs performing better at *LMNB1* (FIG. 31C).

[349] The length of the attB landing site must balance two conflicting factors: the higher efficiency of prime editing for smaller inserts and reduced efficiency of Bxb1 integration at shorter attB lengths. AttB lengths were evaluated at *ACTB, LMNB1,* and nucleolar phosphoprotein p130 *(NOLC1)*, and the optimal attB length was found to be locus dependent. At the *ACTB* locus, long attB lengths could be inserted by prime editing (FIG. 29B) and overall PASTE efficiencies for the insertion of *GFP* were highest for long attB lengths (FIG. 31d). In contrast, intermediate attB lengths had higher overall integration efficiencies (> 20%) at *LMNB1* (FIG. 31E) and *NOLC1* (FIG. 31F), indicating that the increased efficiency of installing shorter attB sequences overcame the reduction of Bxb1 integration at these sites.

[350] The PE3 version of prime editing combines PE2 and an additional nicking guide to bias resolution of the flap intermediate towards insertion. To test the importance of nicking guide selection on PASTE editing, editing at ACTB and LMNB1 loci was tested with two nicking guide positions. Suboptimal nicking guide positions were found to reduce the PASTE efficiency up to 30% (FIG. 32A) in agreement with the 75% reduction of PASTE efficiency in the absence of nicking guide (FIG. 29G). The pegRNA spacer sequence was found to be necessary for PASTE editing, and substitution of the spacer sequence with a non-targeting guide was found to eliminate editing (FIG. 32B).

[351] Rational mutations were also introduced in both the Bxb1 integrase and reverse transcriptase domain of the PE2 construct to optimize PASTE further. While some of these mutations were well tolerated by PASTE (FIGs. 33A-B), none of them improved PASTE editing efficiency.

[352] Short RT and PBS lengths can offer additional improvements for editing. A panel of shorter RT and PBS guides were tested at ACTB and LMNB1 loci and while shorter RT and PBS sequences did not increase editing at ACTB (FIG. 31G), it was found that they had improved editing at LMNB1 (FIG. 31H) with best performing guides reaching GFP insertion rates of ~ 40% (FIG. 31I).

Example 20

PASTE Tagging at Multiple Endogenous Genes

[353] *GFP* insertion efficiency was measured at seven different gene loci – *ACTB, SUPT16H, SRRM2, NOLC1, DEPDC4, NES,* and *LMNB1* – to test the versatility of the

PASTE programming. A range of integration rates up to 22% was found (FIG. 34A). Because PASTE does not require homology or sequence similarity on cargo plasmids, integration of diverse cargo sequences is modular and easily scaled across different loci. Six different gene cargos, varying in size from 969 bp to 4906 bp, were tested for insertion at *ACTB* and *LMNB1* loci with PASTE. Integration frequencies between 5% and 22% depending on the gene and insertion locus were found (FIGs. 34B and 35). Additionally, a panel of seven common therapeutic genes, *CEP290, OTC, HBB, PAH, GBA, BTK,* and *ADA* was evaluated for insertion at the *ACTB* locus, and the efficient integration of these cargos were found between 5%-20% (FIG. 34C).

[354] The precise insertions of PASTE for in-frame protein tagging or expressing cargo without disruption of endogenous gene expression was assessed. As Bxb1 leaves residual sequences in the genome (termed attL and attR) after cargo integration, these genomic scars can serve as protein linkers. The frame of the attR sequence was positioned through strategic placement of the attP on the minicircle cargo, achieving a suitable protein linker, GGLSGQPPRSPSSGSSG (SEQ ID NO: 427). Using this linker, four genes (*ACTB*, *SRRM2*, *NOLC1*, and *LMNB1*) were tagged with *GFP* using PASTE. To assess correct gene tagging, the subcellular location of GFP was compared with the tagged gene product by immunofluorescence. For all four targeted loci, GFP co-localized with the tagged gene product, indicating successful tagging (FIGs. 34D-E).

Example 21

Orthogonal Sequence Preferences for Bxb1 Integration

[355] The central dinucleotide of Bxb1 is involved in the association of attB and attP sites for integration, and changing the matched central dinucleotide sequences can modify integrase activity and provide orthogonality for insertion of two genes. Expanding the set of attB/attP dinucleotides can enable multiplexed gene insertion with PASTE. The efficiency of *GFP* integration at the *ACTB* locus with PASTE across all 16 dinucleotide attB/attP sequence pairs was profiled to find optimal attB/attP dinucleotides for PASTE insertion. Several dinucleotides with integration efficiencies greater than the wild-type GT sequence were found (FIG. 36A). A majority of dinucleotides had 75% editing efficiency or greater compared to wild-type attB/attP efficiency, implying that these dinucleotides can be orthogonal channels for multiplexed gene insertion with PASTE.

[356] The specificity of matched and unmatched attB/attP dinucleotide interactions was then assessed. The interactions between all dinucleotide combinations in a scalable fashion using a pooled assay to compare attB/attP integration were profiled (FIG. 36B). By barcoding 16 attP dinucleotide plasmids with unique identifiers, co-transfecting this attP pool with the Bxb1 integrase expression vector and a single attB dinucleotide acceptor plasmid, and sequencing the resulting integration products, the relative integration efficiencies of all possible attB/attP pairs were measured (FIG. 36C). Dinucleotide specificity was found to vary, with some dinucleotides (GG) exhibiting strong self-interaction with negligible crosstalk, and others (AA) showing minimal self-preference. Sequence logos of attP preferences (FIG. 37) revealed that dinucleotides with C or G in the first position have stronger preferences for attB dinucleotide sequences with shared first bases, while other attP dinucleotides, especially those with an A in the first position, have reduced specificity for the first attB base.

[357] GA, AG, AC, and CT dinucleotide pegRNAs were then tested for *GFP* integration at *ACTB*, either paired with their corresponding attP cargo or mispaired with the other three dinucleotide attP sequences. All four of the tested dinucleotides efficiently were found to integrate cargo only when paired with the corresponding attB/attP pair, with no detectable integration across mispaired combinations (FIG. 36D).

Example 22

Multiplex Gene Integration with PASTE

[358] Multiplexing in cells by using orthogonal pegRNAs that direct a matched attP cargo to a specific site in the genome was assessed (FIG. 38A). Selecting the three top dinucleotide attachment site pairs (CT, AG, and GA), pegRNAs that target *ACTB* (CT), *LMNB1* (AG), and *NOLC1* (GA) and corresponding minicircle cargo containing GFP (CT), mCherry (AG), and YFP (GA) were designed. Upon co-delivering these reagents to cells, single-plex, dual-plex, and trip-plex editing of all possible combinations of these pegRNAs and cargo in the range of 5%-25% integration was found to be achieved (FIG. 38B).

[359] An application for multiplexed gene integration is for labeling different proteins to visualize intracellular localization and interactions within the same cell. PASTE was used to simultaneously tag *ACTB* (GFP) and *NOLC1* (mCherry) or *ACTB* (GFP) and *LMNB1* (mCherry) in the same cell. No overlap of GFP and mCherry fluorescence was observed and tagged genes were confirmed to be visible in their appropriate cellular compartments, based

on the known subcellular localizations of the *ACTB*, *NOLC1* and *LMNB1* protein products (FIGs. 15A-B).

Example 23

PASTE Efficiencies Compared With DSB-based Insertion Methods

[360] PASTE efficiencies were found to exceed comparable DSB-based insertion methods.

[361] PASTE editing was assessed alongside DSB-dependent gene integration using either NHEJ (i.e., homology-independent targeted integration, HITI) or HDR pathways. PASTE had equivalent or better gene insertion efficiencies than either HITI (FIGs. 39A-B) or HDR (FIGs. 39C-D). On a panel of 7 different endogenous targets, PASTE exceeded HITI editing at 6 out of 7 genes, with similar efficiency for the 7th gene (FIG. 39A). As DSB generation can lead to insertions or deletions (indels) as an alternative and undesired editing outcome, the indel frequency of all three methods was assessed by next-generation sequencing, finding significantly fewer indels generated with PASTE than either HDR or HITI in both HEK293FT and HepG2 cells (FIGs. 39B, 39D and 40A), showcasing the high purity of gene integration outcomes with PASTE.

Example 24

Off-target Characterization of PASTE and HITI Gene Integration

[362] Off-target editing can be used in genome editing technologies. The specificity of PASTE at specific sites was assessed based on off-targets generated by Bxb1 integration into pseudo-attB sites in the human genome and off-targets generated via guide- and Cas9-dependent editing in the human genome (FIG. 39E). While Bxb1 lacks documented integration into the human genome at pseudo-attachment sites, potential sites with partial similarity to the natural Bxb1 attB core sequence were computationally identified. Bxb1 integration by ddPCR across these sites was tested and no off-target activity was found (FIGs. 39F and 40B-D). To assay Cas9 off-targets for the *ACTB* pegRNA, two potential off-target sites were identified via computational prediction and no off-target integration for PASTE was found (FIGs. 39G and 40A-D), but substantial off-target activity by HITI at one of the sites was found (FIGs. 39H and 40A-D).

[363] Genome-wide off-targets due to either Cas9 or Bxb1 through tagging and PCR amplification of insert-genomic junctions were additionally assessed (FIG. 39I). Single cell clones were isolated for conditions with PASTE editing and negative controls missing PE2,

and deep sequencing of insert genomic junctions from these clones showed all reads aligning to the on-target *ACTB* site, confirming no off-target genomic insertions (FIGs. 39J-L).

[364] Expression of reverse transcriptases and integrases involved in PASTE can have detrimental effects on cellular health. The complete PASTE system, the corresponding guides and cargo with only PE2, and the corresponding guides and cargo with only Bxb1 were transfected and compared to both GFP control transfections and guides without protein expression via transcriptome-wide RNA sequencing to determine the extent of these effects. While Bxb1 expression in the absence of Prime editing was found to have several significant off targets, the complete PASTE system had only one differentially regulated gene with more than a 1.5-fold change (FIGs. 41A-B). Genes upregulated by Bxb1 overexpression included stress response genes, such as TENT5C and DDIT3, but these changes were not seen in the expression of the PASTE system (FIG. 41C), potentially due to the decreased expression of Bxb1 from the P2A linker on the PASTE construct.

Example 25

PASTE Efficiency in Non-Dividing Cell

[365] PASTE activity in non-dividing cells was assessed. Cas9 and HDR templates or PASTE were transfected into HEK293FT cells and cell division was arrested via aphidicolin treatment (FIG. 42A). In this model of blocked cell division, PASTE was found to maintain a *GFP* gene integration activity greater than 20% at the *ACTB* locus whereas HDR-mediated integration was abolished (FIG. 42B and 43A).

Example 26

Production and Secretion of Therapeutic Transgene

[366] PASTE with larger transgenes and in additional cell lines were assessed.

[367] To evaluate the size limits for therapeutic transgenes, insertion of cargos up to 13.3 kb in length in both dividing and aphidicolin treated cells was assessed. Insertion efficiency greater than 10% was found (FIG. 42C), enabling insertion of ~99.7% of all full-length human cDNA transgenes. To overcome reduction of large insert delivery to cells because of delivery inefficiencies, delivering larger DNA amounts of insert was found to significantly improve gene integration efficiency (FIG. 43B). PASTE editing to additional cell types such as PASTE in the K562 lymphoblast line and in primary human T cells were also assessed. Both PE2-P2A-Bxb1 (PASTE) and separate delivery of PE2 and Bxb1 were found to result in

efficient editing in both cell types (FIGs. 42D-E). Lastly, as therapeutic delivery of PASTE *in vivo* might require viral delivery of the DNA cargo, whether AAV could deliver an attP containing payload that could be integrated into the genome via Bxb1 was evaluated. Targeting the *ACTB* locus, AAV was found to be capable of delivering the appropriate template for integrase mediated insertion with rates up to 4% in a dose dependent fashion (FIGs. 42F and 43C).

[368] To improve the efficiency of PASTE, PE2* NLS was incorporated for prime editing and improved PASTE integration at multiple loci was found (FIG. 44A). Furthermore, PE2* resulted in more robust integration at lower titrations of cargo plasmid, demonstrating integration at amounts as low as 8 ng of plasmid (FIG. 44B). To combat reductions in PASTE efficiency due to incomplete plasmid delivery, a puromycin resistance gene was co-delivered and found to increase the PASTE efficiency in the presence of drug selection (FIG. 45).

[369] Programmable gene integration provides a modality for expression of therapeutic protein products, and protein production was assessed for therapeutically relevant proteins Alpha-1 antitrypsin (encoded by *SERPINA1)* and Carbamoyl phosphate synthetase I (encoded by *CPS1),* involved in the diseases Alpha-1 antitrypsin deficiency and CPS1 deficiency, respectively. By tagging gene products with the luminescent protein subunit HiBiT, the transgene production and secretion were assessed independently in response to PASTE treatment (FIG. 42G). PASTE was transfected with *SERPINA1* or *CPS1* cargo in HEK293FT cells and a human hepatocellular carcinoma cell line (HepG2) and efficient integration at the *ACTB* locus was found (FIG. 42H-I). This integration resulted in robust protein expression, intracellular accumulation of transgene products (FIGs. 42J and 46A-B), and secretion of proteins into the media (FIG. 42K).

Example 27

Optimized PASTE Constructs

[370] To optimize complex activity, a panel of protein modifications were screened, including alternative reverse transcriptase fusions and mutations, various linkers between the reverse transcriptase domain and integrase and between the Cas9 and reverse transcriptase domain, and reverse transcriptase and BxbINT domain mutants (FIG. 47A and FIG. 49C – FIG. 49F). A number of protein modifications, including a 48 residue XTEN linker between the Cas9 and reverse transcriptase and the fusion of MMuLV to the Sto7d DNA binding domain (Oscorbin et al. FEBS Lett. 594. 4338-4356. 2020) improved editing efficiency (FIG. 47A and FIG. 49C – FIG. 49D). When these top modifications were combined with a

GGGGS linker (SEQ ID NO: 420) between the reverse transcriptase-Sto7d domain and the BxbINT, they produced ~55% gene integration, highlighting the importance of directly recruiting the integrase to the target site (FIG. 47A). This optimized construct was referred to as SpCas9-(XTEN-48)-RT-Sto7d-(GGGGS)-BxbINT. The optimized contruct achieved precise integration of templates as large as ~36,000 bp with ~20% integration efficiency (FIG. 47A), with complete integration of the full-length cargo confirmed by Sanger sequencing.

[371] Additionally, pegRNAs containing different AttB length truncations were tested and found that prime editing was capable of inserting sequences up to 56 bp at the beta-actin (*ACTB*) gene locus, with higher efficiency at lengths below 31 bp (FIG. 48A – FIG. 48B). A panel of multiple enzymes was evaluated, including Bxb1 (i.e., BxbINT), TP901 (i.e., Tp9INT), and phiBT1 (i.e., Bt1INT) phage serine integrases. Prime editing successfully inserted all landing sites tested, with efficiencies between 10-30% (FIG. 48C – FIG. 48D)

Example 28

Viral Delivery & In Vivo Editing

[372] In order to package the complete PASTE system in viral vectors, an AdV vector was utilized (FIG. 50B). Adenovirus was evaluated for if it could deliver a suitable template for BxbINT-mediated insertion along with plasmids for SpCas9-RT-BxbINT and guide expression, or AdV delivery of guides and BxbINT with plasmid delivery of SpCas9-RT, finding that 10-20% integration of the ~36 kb adenovirus genome carrying EGFP in HEK293FT and HepG2 cells was achieved (FIG. 50C). Upon packaging and delivering the cargo and PASTE system components across 3 AdV vectors, the complete PASTE system (Cas9-reverse transcriptase, integrase and guide RNAs, or cargo) could be substituted by adenoviral delivery, with integration of up to ~50-60% with viral-only delivery in HEK293FT and HepG2 cells (FIG. 50D).

[373] To further demonstrate PASTE would be amenable for *in vivo* delivery, an mRNA version of the PASTE protein components was developed as well as chemically-modified synthetic atgRNA and nicking guide against the LMNB1 target (FIG. 50E). Electroporation of the mRNA and guides along with delivery of the template via adenovirus or plasmid yielded high efficiency integration up to ~23% (FIG. 50E-FIG. 50F). More sustained BxbINT expression could allow for integration into newly placed AttB sites in the genome, so circular

mRNA expression was tested and found to boost the efficiency of integration to ~30% (FIG. 50G – FIG. 50I).

Example 29

Simultaneous Deletion & Insertion With PASTE

[374] The PASTE system was used to simultaneously delete one sequence and insert another. 130 bp and 385 bp deletions of first exon of LMNB1 with combined insertion of AttB nucleic acid sequence was performed (FIG. 51A). This data shows that it is possible to replace DNA sequence using the PASTE system.

[375] A130 bp deletion of the first exon of LMNB1 with combined insertion of a 967 bp cargo using the PASTE system was also performed.

[376] One of two attP sequences were inserted using the mini circle template that has mutated AttP, as described above. This AttP mutants shows better integration kinetics and efficiency, especially for the shorter AttBs (38-44bp). The LMNB1 AttB used in this experiment is 38 bp (FIG. 51B).

STATEMENTS OF INVENTION

1. A method of site-specific integration of a nucleic acid into a cell genome, the method comprising:

(a) incorporating an integration site at a desired location in the cell genome by introducing into a cell:

i. a DNA binding nuclease linked to a reverse transcriptase domain, wherein the DNA binding nuclease comprises a nickase activity; and

ii. a guide RNA (gRNA) comprising a primer binding sequence linked to an integration sequence, wherein the gRNA interacts with the DNA binding nuclease and targets the desired location in the cell genome, wherein the DNA binding nuclease nicks a strand of the cell genome and the reverse transcriptase domain incorporates the integration sequence of the gRNA into the nicked site, thereby providing the integration site at the desired location of the cell genome; and

(b) integrating the nucleic acid into the cell genome by introducing into the cell:

i. a DNA or RNA strand comprising the nucleic acid linked to a sequence that is complementary or associated to the integration site; and

ii. an integration enzyme, wherein the integration enzyme incorporates the nucleic acid into the cell genome at the integration site by integration, recombination, or reverse transcription of the sequence that is complementary or associated to the integration site, thereby introducing the nucleic acid into the desired location of the cell genome of the cell.

2. The method of claim 1, wherein the gRNA hybridizes to a complementary strand of the cell genome to the genomic strand that is nicked by the DNA binding nuclease.

3. The method of any one of the previous claims, wherein the integration enzyme is introduced as a peptide or a nucleic acid encoding the integration enzyme.

4. The method of any one of the previous claims, wherein the DNA binding nuclease is introduced as a peptide or a nucleic acid encoding the DNA binding nuclease.

5. The method of any one of the previous claims, wherein the DNA or RNA strand comprising the nucleic acid is introduced into the cell as a minicircle, a plasmid, mRNA or a linear DNA.

6. The method of any one of the previous claims, wherein the DNA or RNA strand comprising the nucleic acid is between 1000 bp and 36,000 bp.

7. The method of any one of the claims 1-5, wherein the DNA or RNA strand comprising the nucleic acid is more than 36,000 bp.

8. The method of any one of the claims 1-5, wherein the DNA or RNA strand comprising the nucleic acid is less than 1000 bp.

9. The method of any one of the previous claims, wherein the DNA comprising the nucleic acid is introduced into the cell as a minicircle.

10. The method of claim 9, wherein the minicircle does not comprise a sequence of a bacterial origin.

11. The method of any one of the previous claims, wherein the DNA binding nuclease linked to a reverse transcriptase domain and the integration enzyme are linked via a linker.

12. The method of claim 11, wherein the linker is cleavable.

13. The method of claim 11, wherein the linker is non-cleavable.

14. The method of claim 11, wherein the linker can be replaced by two associating binding domains of the DNA binding nuclease linked to a reverse transcriptase.

15. The method of any one of the previous claims, wherein the integration enzyme is selected from the group consisting of Cre, Dre, Vika, Bxb1, φC31, RDF, FLP, φBT1, R1, R2, R3, R4, R5, TP901-1, A118, φFC1, φC1, MR11, TG1, φ370.1, Wβ, BL3, SPBc, K38, Peaches, Veracruz, Rebeuca, Theia, Benedict, KSSJEB, PattyP, Doom, Scowl, Lockley, Switzer, Bob3, Troube, Abrogate, Anglerfish, Sarfire, SkiPole, ConceptII, Museum, Severus, Airmid, Benedict, Hinder, ICleared, Sheen, Mundrea, BxZ2, φRV, retrotransposases encoded by R2, L1, Tol2 Tc1, Tc3, Mariner (Himar 1), Mariner (mos 1), and Minos, and any mutants thereof.

16. The method of any one of the previous claims, wherein the integration enzyme is Bxb1 or a mutant thereof.

17. The method of any one of the previous claims, wherein the integration site is an attB site, an attP site, an attL site, an attR site, a lox71 site a Vox site, or a FRT site.

18. The method of any one of the previous claims, wherein the DNA binding nuclease comprising a nickase activity is selected from Cas9-D10A, Cas9-H840A, and Cas12a/b nickase.

19. The method of any one of the previous claims, wherein the reverse transcriptase domain is selected from the group consisting of Moloney Murine Leukemia Virus (M-MLV) reverse transcriptase domain, transcription xenopolymerase (RTX), avian myeloblastosis virus reverse transcriptase (AMV-RT), and Eubacterium rectale maturase RT (MarathonRT).

20. The method of any one of the previous claims, wherein the reverse transcriptase domain comprises a mutation relative to the wild-type sequence.

21. The method of claim 19, wherein the M-MLV reverse transcriptase domain comprises one or more mutations selected from the group consisting of D200N, T306K, W313F, T330P and L603W.

22. The method of any one of the previous claims, further comprising introducing a second nicking guide RNA (ngRNA).

23. The method of any one of the previous claims, wherein the gRNA, the nucleic acid encoding the DNA binding nuclease, the reverse transcriptase, the DNA comprising nucleic acid linked to a complementary or associated integration site, the integration enzyme, and optionally the ngRNA, are introduced into a cell in a single reaction.

24. The method of any one of the previous claims, wherein the gRNA, the nucleic acid encoding the DNA binding nuclease, the reverse transcriptase, the DNA comprising nucleic acid linked to a complementary integration site, the integration enzyme, and optionally the ngRNA, are introduced using a virus, a RNP, an mRNA, a lipid, or a polymeric nanoparticle.

25. The method of any one of the previous claims, wherein the nucleic acid is a reporter gene.

26. The method of claim 25, wherein the reporter gene is a fluorescent protein.

27. The method of any one of the previous claims, wherein the cell is a dividing cell.

28. The method of any one of the previous claims, wherein the cell is a non-dividing cell.

29. The method of any one of the previous claims, wherein the desired location in the cell genome is the locus of a mutated gene.

30. The method of any one of the previous claims 1, wherein the nucleic acid is a degradation tag for programmable knockdown of proteins in the presence of small molecules.

31. The method of claim 1, wherein the cell is a mammalian cell, a bacterial cell or a plant cell.

32. The method of claim 1, wherein the nucleic acid is a T-cell receptor (TCR), a chimeric antigen receptor (CAR), an interleukin, a cytokine, or an immune checkpoint gene for integration into a T-cell or natural killer (NK) cell.

33. The method of claim 32, wherein the TCR, the CAR, the interleukin, the cytokine, or the immune checkpoint gene is incorporated into the target site of the T-cell or NK cell genome using a minicircle DNA.

34. The method of claim 1, wherein the nucleic acid is a beta hemoglobin (HBB) gene and the cell is a hematopoietic stem cell (HSC).

35. The method of claim 34, wherein the HBB gene is incorporated into the target site in the HSC genome using a minicircle DNA.

36. The method of claim 34, wherein the nucleic acid is a gene responsible for beta thalassemia or sickle cell anemia.

37. The method of claim 1, wherein the nucleic acid is a metabolic gene.

38. The method of any one of claim 37, wherein the metabolic gene is involved in alpha-1 antitrypsin deficiency or ornithine transcarbamylase (OTC) deficiency.

39. The method of any one of claims 37-38, wherein the metabolic gene is a gene involved in an inherited disease.

40. The method of claim 1, wherein the nucleic acid is a gene involved in an inherited disease or an inherited syndrome.

41. The method of claim 40, wherein the inherited disease is cystic fibrosis, familial hypercholesterolemia, adenosine deaminase (ADA) deficiency, X-linked SCID (X-

SCID), Wiskott-Aldrich syndrome (WAS), hemochromatosis, Tay-Sachs, fragile X syndrome, Huntington's disease, Marfan syndrome, phenylketonuria, or muscular dystrophy.

42. A vector comprising a nucleic acid encoding a DNA binding nuclease comprising a nickase activity C-terminally linked to a reverse transcriptase linked to an integration enzyme via a linker.

43. The vector of claim 42, wherein the linker is cleavable.

44. The vector of claim 42, wherein the linker is non-cleavable.

45. The vector of claim 42, wherein the linker comprises two associating binding domains of the DNA binding nuclease linked to a reverse transcriptase.

46. The vector of claim 42, wherein the integration enzyme comprises a conditional activation domain or conditional expression domain.

47. The vector of claim 46, wherein the integration enzyme is fused to an estrogen receptor.

48. The vector of claim 42, wherein the DNA binding nuclease comprising a nickase activity is selected from the group consisting of Cas9-D10A, Cas9-H840A, and Cas12a/b.

49. The vector of claim 42, wherein the reverse transcriptase is a M-MLV reverse transcriptase, a AMV-RT, MarathonRT, or a RTX.

50. The vector of claim 49, wherein the reverse transcriptase is a modified M-MLV reverse transcriptase relative to the wildtype M-MLV reverse transcriptase.

51. The vector of claim 49, wherein the M-MLV reverse transcriptase domain comprises one or more of the mutations selected from the group consisting of D200N, T306K, W313F, T330P and L603W.

52. The vector of claim 42, wherein the integration enzyme is selected from the group consisting of Cre, Dre, Vika, Bxb1, φC31, RDF, FLP, φBT1, R1, R2, R3, R4, R5, TP901-1, A118, φFC1, φC1, MR11, TG1, φ370.1, Wβ, BL3, SPBc, K38, Peaches, Veracruz, Rebeuca, Theia, KSSJEB, PattyP, Doom, Scowl, Lockley, Switzer, Bob3, Troube, Abrogate, Anglerfish, Sarfire, SkiPole, ConceptII, Museum, Severus, Airmid, Benedict, Hinder, ICleared, Sheen, Mundrea, BxZ2, φRV, retrotransposases encoded

by R2, L1, Tol2 Tc1, Tc3, Mariner (Himar 1), Mariner (mos 1), Minos, and mutants thereof.

53. The vector of claim 52, wherein the recombinase or integrase is Bxb1 or a mutant thereof.

54. A cell comprising:

(a) a vector comprising a nucleic acid encoding a DNA binding nuclease comprising a nickase activity that is C-terminally linked to a reverse transcriptase that is linked to an integration enzyme via a linker;

(b) a gRNA comprising a primer binding sequence, an integration sequence, and a guide sequence, wherein the gRNA can interact with the encoded nuclease comprising a nickase activity;

(c) a DNA minicircle comprising a nucleic acid and a sequence recognized by the encoded integrase, recombinase, or reverse transcriptase; and

(d) a nicking guide RNA (ngRNA) capable of binding the encoded nuclease comprising a nickase activity, wherein the ngRNA targets a sequence away from the gRNA.

55. The cell of claim 54, wherein the minicircle does not comprise a sequence of bacterial origin.

56. The cell of claim 54, wherein the integration enzyme is selected from the group consisting of Cre, Dre, Vika, Bxb1, φC31, RDF, FLP, φBT1, R1, R2, R3, R4, R5, TP901-1, A118, φFC1, φC1, MR11, TG1, φ370.1, Wβ, BL3, SPBc, K38, Peaches, Veracruz, Rebeuca, Theia, KSSJEB, PattyP, Doom, Scowl, Lockley, Switzer, Bob3, Troube, Abrogate, Anglerfish, Sarfire, SkiPole, ConceptII, Museum, Severus, Airmid, Benedict, Hinder, ICleared, Sheen, Mundrea, BxZ2, φRV, retrotransposases encoded by R2, L1, Tol2 Tc1, Tc3, Mariner (Himar 1), Mariner (mos 1), Minos, and any mutants thereof.

57. The cell of claim 54, wherein the integration enzyme is Bxb1 or a mutant thereof.

58. The cell of claim 54, wherein the DNA binding nuclease comprising a nickase activity is selected from the group consisting of Cas9-D10A, Cas9-H840A and Cas12a.

59. The cell of claim 54, wherein the reverse transcriptase is a M-MLV reverse transcriptase.

60. The cell of claim 59, wherein the reverse transcriptase is a modified M-MLV reverse transcriptase.

61. The cell of claim 59, wherein the amino acid sequence of the M-MLV reverse transcriptase comprises one or more mutations selected from the group consisting of D200N, T306K, W313F, T330P, and L603W.

62. The cell of any one of the claims 54-61, further comprising ngRNA.

63. A polypeptide comprising a DNA binding nuclease comprising a nickase activity C-terminally linked to a reverse transcriptase linked to an integration enzyme via a linker.

64. The polypeptide of claim 63, wherein the linker is cleavable.

65. The polypeptide of claim 63, wherein the linker is non-cleavable.

66. The polypeptide of claim 63, wherein the integration enzyme is fused to an estrogen receptor.

67. The polypeptide of claim 63, wherein the DNA binding nuclease comprising a nickase activity is selected from the group consisting of Cas9-D10A, Cas9-H840A, and Cas12a/b/c/d/e/f/g/h/i/j.

68. The polypeptide of claim 63, wherein the reverse transcriptase is a M-MLV reverse transcriptase, a AMV-RT, a MarathonRT, or a XRT.

69. The polypeptide of claim 68, wherein the reverse transcriptase is a modified M-MLV relative to a wild-type M-MLV reverse transcriptase.

70. The polypeptide of claim 69, wherein the M-MLV reverse transcriptase domain comprises one or more of mutations selected from the group consisting of D200N, T306K, W313F, T330P, and L603W.

71. The polypeptide of claim 63, wherein the integration enzyme is selected from group consisting of Cre, Dre, Vika, Bxb1, φC31, RDF, FLP, φBT1, R1, R2, R3, R4, R5, TP901-1, A118, φFC1, φC1, MR11, TG1, φ370.1, Wβ, BL3, SPBc, K38, Peaches, Veracruz, Rebeuca, Theia, KSSJEB, PattyP, Doom, Scowl, Lockley, Switzer, Bob3, Troube, Abrogate, Anglerfish, Sarfire, SkiPole, ConceptII, Museum, Severus, Airmid, Benedict, Hinder, ICleared, Sheen, Mundrea, BxZ2, φRV, retrotransposases encoded

by R2, L1, Tol2 Tc1, Tc3, Mariner (Himar 1), Mariner (mos 1), Minos, and any mutants thereof.

72. The polypeptide of claim 63, wherein the integration enzyme is Bxb1 or a mutant thereof.

73. A gRNA that specifically binds to a DNA binding nuclease comprising nickase activity, the gRNA comprising:

(a) a primer binding site, which hybridizes to a nicked DNA strand;

(b) a recognition site for an integration enzyme; and

(c) a target recognition sequence recognizing a target site in a cell genome and hybridizing to a genomic strand complementary to the strand that is nicked by the DNA binding nuclease.

74. The gRNA of claim 73, wherein the DNA binding nuclease comprising a nickase activity is selected from the group consisting of Cas9-D10A, Cas9-H840A, and Cas12a/b/c/d/e/f/g/h/i/j.

75. The gRNA of claim 73, wherein the primer binding site hybridizes to the 3′ end of the nicked DNA strand.

76. The gRNA of claim 73, wherein the recognition site for the integration enzyme is selected from an attB site, an attP site, an attL site, an attR site, a lox71 site, and a FRT site.

77. The gRNA of claim 73, wherein the recognition site for the integration enzyme is a Bxb1 site.

78. A method of site-specific integration of two or more nucleic acids into a cell genome, the method comprising:

(a) incorporating two integration sites at desired locations in the cell genome by introducing into the cell:

i. a DNA binding nuclease linked to a reverse transcriptase domain, wherein the DNA binding nuclease comprises a nickase activity; and

ii. two guide RNAs (gRNAs), each comprising, a primer binding sequence, and is linked to a unique integration sequence, wherein the gRNA interacts with the DNA binding nuclease and targets the desired locations in the cell genome, wherein the DNA binding nuclease nicks a strand of the cell genome and the

reverse transcriptase domain incorporates each of the integration sequence of the gRNA into the nicked site, thereby providing the integration site at the desired locations of the cell genome; and

(b) integrating the nucleic acid by introducing into the cell:

i. two or more DNA or RNA comprising the nucleic acids, wherein each DNA is flanked by orthogonal integration sites; and

ii. an integration enzyme, wherein the integration enzyme incorporates the nucleic acids into the cell genome at the integration sites by integrase, recombinase, or reverse transcriptase of the sequence that is complementary or associated to the integration site, thereby introducing the nucleic acids into the desired locations of the cell genome of the cell.

79. The method of claim 78, wherein each of the two different integration sites inserted into the cell genome are attB sequences comprising different palindromic or non-palindromic central dinucleotide.

80. The method of claim 78, wherein each of the two different integration sites inserted into the cell genome are attP sequences comprising different palindromic or non-palindromic central dinucleotide.

81. The method of claim 78, wherein the integration enzyme enables each of the two or more DNA or RNA comprising the nucleic acids to directionally enable integration of the nucleic acids into a genome via recombination of a pair of orthogonal attB site sequence and an attP site sequence.

82. The method of claim 78, wherein the integration enzyme is selected from the group consisting of Cre, Dre, Vika, Bxb1, φC31, RDF, FLP, φBT1, TP901-1, A118, φFC1, φC1, MR11, TG1, φ370.1, Wβ, BL3, SPBc, K38, Peaches, Veracruz, Rebeuca, Theia, KSSJEB, PattyP, Doom, Scowl, Lockley, Switzer, Bob3, Troube, Abrogate, Anglerfish, Sarfire, SkiPole, ConceptII, Museum, Severus, Airmid, Benedict, Hinder, ICleared, Sheen, Mundrea, BxZ2, φRV, retrotransposases encoded by R1, R2, R3, R4, R5, L1, Tol2 Tc1, Tc3, Mariner (Himar 1), Mariner (mos 1), Minos, and any mutants thereof.

83. The method of claim 78, wherein the integration enzyme is Bxb1 or a mutant thereof.

84. The method of claim 78, wherein the DNA comprising genes are genes involved in a cell maintenance pathway, cell-division, or a signal transduction pathway.

85. The method of claim 78, wherein the reverse transcriptase domain comprises Moloney Murine Leukemia Virus (M-MLV) reverse transcriptase domain, transcription xenopolymerase (RTX), or avian myeloblastosis virus reverse transcriptase (AMV-RT).

86. The method of claim 78, wherein the DNA binding nuclease comprising a nickase activity is selected from the group consisting of Cas9-D10A, Cas9-H840A, and Cas12a/b.

87. The method of claim 78, wherein the pair of an attB site sequence and an attP site sequence are selected from the group consisting of SEQ ID NO: 5 and SEQ ID NO: 6, SEQ ID NO: 7 and SEQ ID NO: 8, SEQ ID NO: 9 and SEQ ID NO: 10, SEQ ID NO: 11 and SEQ ID NO: 12, SEQ ID NO: 13 and SEQ ID NO: 14, SEQ ID NO: 15 and SEQ ID NO: 16, SEQ ID NO: 17 and SEQ ID NO: 18, SEQ ID NO: 19 and SEQ ID NO: 20, SEQ ID NO: 21 and SEQ ID NO: 22, SEQ ID NO: 23 and SEQ ID NO: 24, SEQ ID NO: 25 and SEQ ID NO: 26, SEQ ID NO: 27 and SEQ ID NO: 28, SEQ ID NO: 29 and SEQ ID NO: 30, SEQ ID NO: 31 and SEQ ID NO: 32, SEQ ID NO: 33 and SEQ ID NO: 34, and SEQ ID NO: 35 and SEQ ID NO: 36.

88. A cell comprising:

(a) a vector comprising a nucleic acid encoding a DNA binding nuclease comprising a nickase activity, wherein the DNA binding nuclease is C-terminally linked to a reverse transcriptase, wherein the reverse transcriptase is linked to a recombinase or integrase via a linker;

(b) two guide RNAs (gRNAs) comprising a primer binding sequence, an integration sequence and a guide sequence, wherein the gRNA can interact with the encoded DNA binding nuclease comprising a nickase activity;

(c) two or more DNA or RNA strands comprising a nucleic acid and a pair of flanking attB site sequence and an attP site sequence recognized by the encoded integrase or recombinase; and optionally

(d) a nicking guide RNA (ngRNA) capable of binding the encoded nuclease comprising a nickase activity, and wherein the ngRNA targets a sequence away from the gRNA.

89. A cell comprising a modified genome, wherein the modification comprises incorporation of two orthogonal integration sites within a cell genome by introducing into the cell:

(a) a vector comprising a nucleic acid encoding a DNA binding nuclease comprising a nickase activity, wherein the DNA binding nuclease is C-terminally linked to a reverse transcriptase;

(b) two guide RNAs (gRNAs), each comprising a primer binding sequence, a genomic integration sequence, and a guide sequence, wherein the gRNA can interact with the encoded nuclease comprising a nickase activity; and optionally,

(c) a nicking guide RNA (ngRNA) capable of binding the encoded nuclease comprising a nickase activity, wherein the ngRNA targets a sequence away from the gRNA.

90. A method of integrating two or more nucleic acids into the cell genome of cell of claim 90, the method comprising introducing into the cell:

(a) two or more DNA, each comprising a nucleic acid and a pair of flanking orthogonal integration site sequences;

(b) an integration enzyme that can recognize the integration site sequence enabling directional linking of the two or more DNA comprising nucleic acid; and

(c) enabling incorporation of the nucleic acids into the cell genome by integrating the 5' orthogonal integration sequence of a first DNA with the first genomic integration sequence and 3' orthogonal integration sequence of a last DNA with the last genomic integration sequence, thereby incorporating the two or more nucleic acids into the cell genome.

91. A cell comprising a modified genome, wherein a modification of the modified genome comprises incorporation of two orthogonal integration sites within the cell genome by introducing into the cell:

(a) a vector comprising a nucleic acid encoding a DNA binding nuclease comprising a nickase activity, wherein the DNA binding nuclease is C-terminally linked to a reverse transcriptase;

(b) two guide RNAs (gRNAs), each comprising a primer binding sequence, a genomic integration sequence, and a guide sequence, wherein the gRNA can interact with the encoded nuclease comprising a nickase activity; and optionally

(c) a nicking guide RNA (ngRNA) capable of binding the encoded nuclease comprising a nickase activity, and wherein the ngRNA targets a sequence away from the gRNA;

(d) two or more DNA or RNA comprising the nucleic acids, wherein each DNA is flanked by orthogonal integration sites; and

(e) an integration enzyme, wherein the integration enzyme incorporates the nucleic acids into the cell genome at the integration sites.

CLAIMS

1. A system capable of site-specifically integrating an exogenous nucleic acid into a mammalian cell genome at a desired target site, wherein the system comprises, in a single composition:

(a) a nucleic acid encoding a DNA binding nickase domain linked to a reverse transcriptase domain;

(b) a nucleic acid encoding at least two guide RNAs (gRNAs) comprising, from 3' to 5',

i. a primer binding sequence,

ii. a sequence complementary to one strand of an integration recognition sequence, and

iii. a target binding sequence,

wherein the gRNA is capable of guiding the linked nickase-reverse transcriptase domains to the genomic target site;

(c) nucleic acid encoding an integration enzyme; and

(d) an exogenous nucleic acid linked to a sequence that is an integration cognate of the integration recognition sequence.

2. The system of claim 1, wherein the DNA binding nickase domain is linked to the reverse transcriptase domain by in-frame fusion.

3. The system of claim 1, wherein the DNA binding nickase domain is linked to the reverse transcriptase domain by a linker.

4. The system of claim 3, wherein the linker is a peptide fused in-frame between the nickase and reverse transcriptase domains.

5. The system of claim 1, wherein the linked DNA binding nickase-reverse transcriptase domains are further linked to the integration enzyme.

6. The system of claim 1, wherein the DNA binding nickase domain is

selected from Cas9-D10A, Cas9-H840A, and Cas12a/b nickase.

7. The system of claim 1, wherein the reverse transcriptase domain is selected from the group consisting of Moloney Murine Leukemia Virus (M-MLV) reverse transcriptase domain, transcription xenopolymerase (RTX), avian myeloblastosis virus reverse transcriptase (AMV-RT), and Eubacterium rectale maturase RT.

8. The system of claim 7, wherein the reverse transcriptase domain is a M-MLV reverse transcriptase domain.

9. The system of claim 7, wherein the M-MLV reverse transcriptase domain comprises one or more mutations selected from the group consisting of D200N, T306K, W313F, T330P, and L603W.

10. The system of claim 1, wherein the exogenous nucleic acid is a minicircle, a plasmid, a mRNA, or a linear DNA

11. The system of claim 9, wherein exogenous nucleic acid is a minicircle.

12. The system of claim 11, wherein the minicircle does not comprise a sequence of a bacterial origin.

13. The system of claim 1, wherein the integration enzyme is selected from the group consisting of Dre, Vika, Bxb1, φC31, RDF, FLP, φBT1, R1, R2, R3, R4, RS, TP901-1, A118, φFC1, φC1, MR11, TG1, φ370.1, Wβ, BL3, SPBc, K38, Peaches, Veracruz, Rebeuca, Theia, Benedict, KSSJEB, PattyP, Doom, Scowl, Lockley, Switzer, Bob3, Troube, Abrogate, Anglerfish, Sarfire, SkiPole, ConceptII, Museum, Severus, Airmid, Benedict, Hinder, ICleared, Sheen, Mundrea, BxZ2, φRV, retrotransposases encoded by R2, L1, Tol2 Te1, Tc3, Mariner Rimar 1, Mariner mos-1, and Minos.

14. The system of claim 13, wherein the integration enzyme is Bxb1.

15. The system of claim 13, wherein the integration recognition sequence is an attB sequence, an attP sequence, a Vox sequence, or a FRT sequence.

16. The system of claim 15, wherein the integration recognition sequence is an attB sequence and the integration cognate is an attP sequence.

17. The system of claim 15, wherein the exogenous nucleic acid encodes: a reporter gene;
a degradation tag for programmable knockdown of proteins in the presence of small molecules;
a T-cell receptor (TCR), a chimeric antigen receptor (CAR), an interleukin, a cytokine, or an immune checkpoint gene and the mammalian cell is a T-cell or natural killer (NK) cell;
a beta hemoglobin (HBB) gene and the mammalian cell is a hematopoietic stem cell (HSC);
a metabolic gene; or
a gene involved in an inherited disease or syndrome.

18. The system of claim 1, wherein the exogenous nucleic acid is between 1000 bp and 36,000 bp in length.

19. The system of claim 1, wherein the exogenous nucleic acid is more than 36,000 bp in length.

20. The system of claim 1, wherein the exogenous nucleic acid is less than 1000 bp in length.

21. The system of claim 17, wherein the inherited disease is cystic fibrosis, familial hypercholesterolemia, adenosine deaminase (ADA) deficiency, X-linked SCID (X-SCID), Wiskott-Aldrich syndrome (WAS), hemochromatosis, Tay-Sachs, fragile X syndrome, Huntington's disease, Marfan syndrome, phenylketonuria, or muscular dystrophy.

22. The system of claim 1, further comprising a nicking gRNA.

23. The system of claim 1, wherein a)-d) are introduced into the mammalian cell as an adeno-associated virus (AAV) or an adenovirus (AdV).

24. The system of claim 1, wherein the nucleic acid encoding at least two gRNAs comprises a nucleic acid encoding two gRNAs.

25. The system of claim 1, wherein the composition is *ex vivo* or *in vitro*.