

ENGINEERED RETROTRANSPOSABLE ELEMENT PROTEINS FOR GENE DELIVERY AND INSERTION

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims priority to U.S. Provisional Application No. 63/588,418 filed on October 6, 2023, U.S. Provisional Application No. 63/558,395, filed February 27, 2024, and U.S. Provisional Application No. 63/640,141, filed on April 29, 2024. These applications are hereby incorporated by reference in their entirety.

REFERENCE TO AN ELECTRONIC SEQUENCE LISTING

[0002] The contents of the electronic sequence listing (EXXS_004_03WO_SeqList_ST26.xml; Size: 18,172,381 bytes; and Date of Creation: October 3, 2024) are herein incorporated by reference in its entirety.

BACKGROUND

[0003] Gene therapy can be used to treat diseases or conditions associated with genetic defects by delivering one or more therapeutic genes or cells with corrected defects to a subject. For example, therapeutic genes can be delivered in vectors that promote integration into the host genome of a subject.

[0004] Despite recent activity in the field of retroelement delivery and genome insertion, there remains a need for nucleic acids and proteins to carry out these methods with improved, inter alia, efficiency, specificity, accuracy, fidelity and processivity. Provided herein are compositions and methods that address this need.

SUMMARY

[0005] According to one aspect, the present disclosure relates to a nucleic acid encoding an engineered driver comprising an R2 protein or portion thereof fused to at least one heterologous polypeptide, wherein: the at least one heterologous polypeptide comprises a heterologous nucleosome binding polypeptide, a heterologous nucleic acid binding polypeptide, a heterologous RNA/DNA repair polypeptide, a heterologous RNA/DNA processing polypeptide,

a heterologous localization signal, a heterologous reverse transcriptase, a heterologous endonuclease, a heterologous intrinsically disordered polypeptide, or any combination thereof; and the engineered driver exhibits an improvement in at least one of efficiency, specificity, accuracy, fidelity, or processivity compared to a reference R2 protein. In certain embodiments, the engineered driver exhibits an improvement in at least efficiency compared to a reference R2 protein

[0006] According to another aspect, the present disclosure relates to a nucleic acid encoding an engineered driver comprising an R2 protein or portion thereof, a heterologous CTIP polypeptide, and a heterologous HMGN1 polypeptide.

[0007] According to another aspect, the present disclosure relates to a nucleic acid encoding an engineered driver comprising an R2 protein or portion thereof, a heterologous Sto7d polypeptide, and a heterologous HMGN1 polypeptide.

[0008] According to another aspect, the present disclosure relates to a nucleic acid encoding an engineered driver comprising an amino acid sequence with one or more modifications relative to any one of SEQ ID NOS: 30001-30060. In some embodiments, the engineered driver comprises an amino acid sequence set forth in SEQ ID NOS: 30001-30060, or an amino acid sequence that is at least 70% identical thereto. In some embodiments, the engineered driver comprises a reverse transcriptase domain and an endonuclease domain.

[0009] According to another aspect, the present disclosure relates to a nucleic acid encoding an engineered driver comprising a R2 protein or portion thereof, wherein the R2 protein or portion comprises the amino acid substitution of D555K relative to SEQ ID NO:1.

[0010] According to another aspect, the present disclosure relates to a nucleic acid encoding an engineered driver comprising an RTE polypeptide fused to a heterologous intrinsically disordered polypeptide.

[0011] According to another aspect, the present disclosure relates to a nucleic acid encoding an engineered driver comprising an RTE polypeptide fused to a RNA helicase polypeptide.

[0012] According to another aspect, the present disclosure relates to an engineered driver encoded by a nucleic acid of the disclosure.

[0013] According to another aspect, the present disclosure relates to a transgene expression system comprising: a driver construct comprising the nucleic acid of any one claims 1-140, or the engineered driver of claim 141; and a gene delivery construct comprising a transgene and at least one retroelement untranslated region (RTE-UTR) derived from a RTE, wherein the at least

one RTE-UTR is capable of binding the engineered driver, and the engineered driver is configured to mediate integration of the transgene into a genome. In certain embodiments, the RTE-UTR is derived from a R2 RTE.

[0014] According to another aspect, the present disclosure relates to a gene delivery construct comprising: a transgene; and at least one R2 retroelement untranslated region (RTE-UTR) derived from a R2 RTE, wherein the RTE-UTR is capable of binding a R2 protein, and wherein: the at least one RTE-UTR comprises a 3' RTE-UTR, a 5' RTE-UTR, or a combination thereof, the 3' RTE-UTR has a nucleic acid sequence set forth in any one of SEQ ID NOS: 1338-1340, 1344-1347, 1352-1354, 1843, 1865, and 33001-33060, and the 5' RTE-UTR has a nucleic acid sequence set forth in any one of SEQ ID NOS: 1341-1343, 1348-1351, 1355-1360-1365, 1366, 1838-1850, 1873-1874, 1862-1863, 1866-1872, and 35001-35060.

[0015] According to another aspect, the present disclosure relates to a lipid nanoparticle (LNP) comprising: a nucleic acid of the disclosure; an engineered driver of the disclosure; a transgene expression system of the disclosure; a gene delivery construct of the disclosure; or a combination thereof.

[0016] According to another aspect, the present disclosure relates to a pharmaceutical composition comprising a pharmaceutically acceptable carrier and: a nucleic acid of the disclosure; an engineered driver of the disclosure; a transgene expression system of the disclosure; a gene delivery construct of the disclosure; an LNP of the disclosure; or a combination thereof.

[0017] According to another aspect, the present disclosure relates to a method of promoting integration of a transgene into a genome of a cell, the method comprising contacting a cell with: a nucleic acid of the disclosure; an engineered driver of the disclosure; a transgene expression system of the disclosure; a gene delivery construct of the disclosure; an LNP of the disclosure; a pharmaceutical composition of the disclosure; or a combination thereof.

[0018] According to another aspect, the present disclosure relates to a method of treating a disease in a subject in need thereof, comprising administering to the subject a therapeutically effective amount of: a nucleic acid of the disclosure; an engineered driver of the disclosure; a transgene expression system of the disclosure; a gene delivery construct of the disclosure; an LNP of the disclosure; a pharmaceutical composition of the disclosure; or a combination thereof.

[0019] According to another aspect, the present disclosure relates to a nucleic acid of the disclosure; an engineered driver of the disclosure, a transgene expression system of the

disclosure, a gene delivery construct of the disclosure, an LNP of the disclosure, a pharmaceutical composition of the disclosure, or a combination thereof, for use in a method of treating the human or animal body by therapy.

[0020] According to another aspect, the present disclosure relates to a use of a nucleic acid of the disclosure, an engineered driver of the disclosure, a transgene expression system of the disclosure, a gene delivery construct of the disclosure, an LNP of the disclosure, a pharmaceutical composition of the disclosure, or a combination thereof, for the manufacture of a medicament.

BRIEF DESCRIPTION OF THE DRAWINGS

[0021] The figures and figure descriptions provided herein are intended to illustrate embodiments by way of example only.

[0022] FIGS. 1A-1E illustrate non-limiting examples of nucleic acids that are useful for promoting insertion of a transgene (e.g., a heterologous nucleic acid comprising a gene of interest) into a target nucleic acid.

[0023] FIG. 1A illustrates a non-limiting example of a gene delivery construct (gene delivery nucleic acid) comprising a transgene flanked by two terminal regions that can interact with a driver and promote integration of the transgene into a target nucleic acid (e.g., genomic DNA).

[0024] FIG. 1B illustrates a non-limiting example of two separate nucleic acid molecules, wherein the first nucleic acid is a gene delivery construct comprising a gene of interest flanked by two terminal regions, and the second nucleic acid encodes an engineered driver comprising a retrotransposable element (RTE) polypeptide. Such nucleic acids are “in trans” or as “trans” nucleic acids or constructs.

[0025] FIG. 1C illustrates a non-limiting example of a single nucleic acid comprising both a gene of interest and a sequence encoding an engineered driver comprising a RTE polypeptide, flanked by two terminal regions. Such a nucleic acid is known as an “in cis” nucleic acid or construct.

[0026] FIG. 1D illustrates a non-limiting example of gene integration into a target nucleic acid (e.g., genomic DNA) promoted by two separate nucleic acid molecules (in trans; a gene delivery

nucleic acid comprising a gene of interest flanked by two terminal regions, and a driver nucleic acid comprising a coding sequence for an engineered driver.

[0027] FIG. 1E illustrates a non-limiting example of gene integration into a target nucleic acid (e.g., genomic DNA) promoted by a single nucleic acid comprising both a gene of interest and an engineered driver coding sequence flanked by two terminal regions, in cis.

[0028] FIGS. 2A-2L illustrate non-limiting examples of different configurations of an engineered driver comprising a RTE polypeptide fused to one or more heterologous polypeptides that can promote integration of a transgene into a target nucleic acid.

[0029] FIG. 3 illustrates three non-limiting aspects of an ORF1-like RNA-binding protein (ORBP) system. The ORBP system provides one or more retroelement proteins and one or more polypeptide comprising RNA binding activity.

[0030] FIGS. 4A, 4B and 4C are bar graphs showing results of integration assays using drivers comprising polypeptide fusions and mutations, amino acid substitutions and deletions in R2-TG derived retroelements and reporters in cis.

[0031] FIG. 5 is a bar graph showing results of integration assays using drivers comprising polypeptide fusions and mutations in R2-ZA derived retroelements and reporters in cis.

[0032] FIG. 6 is a bar graph showing results of integration assays using drivers comprising polypeptide fusions and mutations in R2-ZA derived retroelements and reporters in cis.

[0033] FIG. 7 is a bar graph showing results of integration assays using drivers comprising polypeptide fusions and mutations in R4-AC derived retroelements and reporters in cis.

[0034] FIG. 8 is a bar graph showing results of integration assays using R2-TG derived retroelements and GFP reporters containing pseudoknots compared to one with a polyA tail.

[0035] FIGS. 9A and 9B are bar graphs showing results of integration assays in primary hepatocytes in donor 1 (9A) or donor 2 (9B).

[0036] FIGS. 10A-10B demonstrate GFP transgene integration with LNPs (non-targeted) containing mRNA for R2-1_TG engineered driver with different MNDopt-GFP R2 reporters and different ratios 2 days post TransAct activation.

[0037] FIGS. 11A and 11B are bar graphs showing results of an integration assay in U2OS cells transfected with GFP reporters with RTE-UTRs created through ancestral reconstruction and an engineered R2-TG driver.

[0038] FIGS. 12A and 12B are bar graphs showing results of an integration assay in HUH7 cells transfected with GFP reporters with modified RTE-UTRs and an engineered R2-TG driver.

[0039] FIGS. 13A and 13B are bar graphs showing results of an integration assay in HUH7 cells transfected with GFP reporters with multiple copies of RTE-UTRs and an engineered R2-TG driver.

[0040] FIGS. 14A and 14B are bar graphs showing the effect of R2-TG drivers engineered with combinations of heterologous polypeptide fusions and amino acid substitutions on integration of a transgene in HUH7 cells.

[0041] FIGS. 15A and 15B are bar graphs showing the effect of R2-TG drivers engineered with combinations of N-terminal truncation, heterologous polypeptide fusions and amino acid substitutions on integration of a transgene in HUH7 cells.

[0042] FIGS. 16A and 16B are bar graphs showing the effect of R2-TG drivers engineered with combinations of N-terminal truncation, heterologous polypeptide fusions and amino acid substitutions on integration of a transgene in HUH7 cells.

[0043] FIGS. 17A and 17B are bar graphs showing the effect of R2-TG drivers engineered with combinations of N-terminal truncation, heterologous polypeptide fusions and amino acid substitutions on integration of a transgene in HUH7 cells.

[0044] FIGS. 18A and 18B are bar graphs showing the effect of R2-TG drivers engineered with combinations of N-terminal truncation, heterologous polypeptide fusions and amino acid substitutions on integration of a transgene in HUH7 cells.

[0045] FIGS. 19A and 19B are bar graphs showing the effect of R2-TG drivers engineered with combinations of N-terminal truncation, heterologous polypeptide fusions and amino acid substitutions on integration of a transgene in HUH7 cells.

[0046] FIGS. 20A and 20B are bar graphs showing the effect of R2-TG drivers engineered with inter alia combinations of N-terminal truncation, RNA helicase fusions and D923K amino acid substitutions on integration of a transgene in HUH7 cells

[0047] FIGS. 21A and 21B are bar graphs showing the effect of R2-TG drivers engineered with inter alia combinations of N-terminal truncation, condensation polypeptide fusions and D923K amino acid substitutions on integration of a transgene in HUH7 cells.

[0048] FIGS. 22A and 22B are bar graphs showing the effect of R2-TG drivers engineered with inter alia combinations of N-terminal truncation, condensation polypeptide fusions and D923K amino acid substitution on integration of a transgene in HUH7 cells.

[0049] FIGS. 23A and 23B are bar graphs showing the effect of R2-TG drivers engineered with inter alia combinations of N-terminal truncation, modified CtIP polypeptide fusions and amino acid substitutions on integration of a transgene in HUH7 cells.

[0050] FIGS. 24A and 24B are bar graphs showing the effect of R2-TG drivers engineered with inter alia combinations of N-terminal truncation, modified CtIP polypeptide fusions and amino acid substitutions on integration of a transgene in HUH7 cells.

[0051] FIGS. 25A and 25B are bar graphs showing the effect of engineered R2-TG drivers and R2 chimeric drivers on integration of a transgene in HUH7 cells.

[0052] FIGS. 26A and 26B are bar graphs showing the effect of R2-TG drivers engineered with inter alia combinations of N-terminal truncation, condensation polypeptide fusions or RNA helicase fusions and amino acid substitutions on integration of a transgene in HUH7 cells.

[0053] FIGS. 27A and 27B are bar graphs showing the effect of R2-TG drivers engineered with inter alia combinations of N-terminal truncation, heterologous polypeptide fusions and amino acid substitutions on integration of a transgene in HUH7 cells.

[0054] FIGS. 28A and 28B are bar graphs showing the effect of R2-TG drivers engineered using the driver having amino acid sequence set forth in SEQ ID NO: 715 with a D932K amino acid substitution as template for the constructs on integration of a transgene in HUH7 cells.

[0055] FIGS. 29A and 29B are bar graphs showing the effect of R2-TG drivers engineered using the driver having amino acid sequence set forth in SEQ ID NO: 715 with a D932K amino acid substitution as template for the constructs on integration of a transgene in HUH7 cells.

[0056] FIGS. 30A and 30B are bar graphs showing the effect of R2-TG drivers engineered using the driver having amino acid sequence set forth in SEQ ID NO: 715 with a D932K amino acid substitution as template for the constructs on integration of a transgene in HUH7 cells.

[0057] FIGS. 31A and 31B are bar graphs showing results of integration assays using engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOs: 853-868.

[0058] FIGS. 32A and 32B are bar graphs showing results of integration assays using engineered R2-TG driver having amino acid sequences set forth in SEQ ID NOs: 869-883.

[0059] FIGS. 33A and 33B are bar graphs showing results of integration assays using engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOs: 884-899.

[0060] FIGS. 34A and 34B are bar graphs showing results of integration assays using engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOs: 900-915.

[0061] FIGS. 35A and 35B are bar graphs showing results of integration assays using engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOs: 916-931.

[0062] FIGS. 36A and 36B are bar graphs showing results of integration assays using engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOs: 932-948.

[0063] FIGS. 37A and 37B are bar graphs showing results of integration assays using engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOs: 949-963.

[0064] FIGS. 38A and 38B are bar graphs showing results of integration assays using engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOs: 884, 964-972 and 974-980.

[0065] FIGS. 39A and 39B are bar graphs showing results of integration assays using engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOs: 981-996.

[0066] FIGS. 40A and 40B are bar graphs showing results of integration assays using engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOs: 997-1012.

[0067] FIGS. 41A and 41B are bar graphs showing results of integration assays using engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOs: 1013-1026.

[0068] FIGS. 42A and 42B are bar graphs showing results of integration assays using engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOs: 1028-1044.

[0069] FIGS. 43A and 43B are bar graphs showing results of integration assays using engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOs: 1045-1061.

[0070] FIGS. 44A and 44B are bar graphs showing results of integration assays using engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOs: 1062-1076.

[0071] FIGS. 45A and 45B are bar graphs showing results of integration assays using engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOs: 1077-1092.

[0072] FIGS. 46A and 46B are bar graphs showing results of integration assays using engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOs: 1093-1107.

[0073] FIGS. 47A and 47B are bar graphs showing results of integration assays using engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOs: 1109-1124.

[0074] FIGS. 48A and 48B are bar graphs showing results of integration assays using engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOs: 1125-1136.

[0075] FIGS. 49A and 49B are bar graphs showing results of integration assays using engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOs: 1137-1148.

[0076] FIGS. 50A and 50B are bar graphs showing results of integration assays using engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOs: 1149-1159 and 161-162.

[0077] FIGS. 51A and 51B are bar graphs showing results of integration assays using engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOs: 1163-1170 and 1977-1978.

[0078] FIGS. 52A and 52B are bar graphs showing results of integration assays using engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOs: 764, 765, 770-774, 1321, 1322 and 1845 and one R2 reporter constructed by ancestor reconstruction process set forth in SEQ ID NO: 1323.

[0079] FIGS. 53A and 53B are bar graphs showing results of integration assays using engineered R2-TG drivers comprising the combinations of retrotransposable elements R2-TG, R2-PTT, R2-LS, R2-GA and chimeras generated from these R2 RTEs.

[0080] FIGS. 54A and 54 B are bar graphs showing results of integration assays using engineered R2-TG drivers comprising the different retrotransposable chimera elements composed from R2-TG, R2-PTT, R2-LS, R2-GA (SEQ ID NO:1293, SEQ ID NO:1417, SEQ ID NO:1420 and SEQ ID NOs: 1335, 471, 410, 408).

[0081] FIGS. 55 is bar graphs showing results of integration assays using engineered R2-TG drivers comprising the different retrotransposable chimera elements (SEQ ID NOS 1305-1308) composed from R2-TG, R2-PTT, R2-LS, R2-GA (SEQ ID NO:1293, SEQ ID NO:1417, SEQ ID NO:1420 SEQ ID NOs: 1335, 471, 410, 408).

[0082] FIGS. 56A and 56B are bar graphs showing results of integration assays using RTE drivers created by ancestor reconstruction process on R2-TG variants, and a chimera built from active R2 elements in addition to transgene cassette reporters for R2-PTT, R2-LS and R2-GFo and reporters constructed UTRs from R2-TG variants.

[0083] FIGS. 57A and 57B are bar graphs showing results of integration assays using chimeric RTE drivers composed from R2-TG, R2-LS, R2-GP and R2-MC (SEQ ID NOs: 1335, 636, 410, 453) and transgene cassette reporter with 3'UTR created by ancestral reconstruction process using 3'UTRs from active R2 elements (SEQ ID NO:1873).

[0084] FIGS. 58A and 58B are bar graphs showing results of integration assays using chimeric RTE drivers composed from R2-TG, R2-PTT, R2-GP and R2-XC SEQ ID NOs: 636, 1417, 453, 474).

[0085] FIG. 59 is a bar graph showing results of integration assays using chimeric RTE drivers composed from R2-TG, R2-PTT, R2-LS, R2-MC and R2-XC species and engineered R2-TG driver SEQ ID NO:636.

[0086] FIGS. 60A and 60B are bar graphs showing results of integration assays using chimeric RTE drivers composed from R2-TG, R2-PTT, R2-MC, R2-GP and R2-XC set forth in SEQ ID NOs: 1335, 636, 410, 458, 453, 474.

[0087] FIGS. 61A and 61B are bar graphs showing results of integration assays using chimeric RTE drivers composed from R2-TG, R2-PTT, R2-LS and R2-GA.

[0088] FIGS. 62A and 62B are bar graphs showing results of integration assays using engineered R2-TG drivers amino acid substitution having an amino acid sequence set forth in SEQ ID NOs: SEQ ID NOS: 1293, 1375, 1390, 1397, 1398, 1422, 1596, 1597, 1636, 1712, 1725-1734.

[0089] FIGS. 63A and 63B are bar graphs showing results of integration assays using engineered R2-TG drivers amino acid substitution having an amino acid sequence set forth in SEQ ID NOs: 1473, 1477, 1519 and 1735-1745.

[0090] FIGS. 64A and 64B are bar graphs showing results of integration assays using engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOs: 1390, 1743, and 1750-1764.

[0091] FIGS. 65A and 65B are bar graph showing results of integration assays using engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOs: 1781-1786 and 1788-1793.

[0092] FIGS. 66A and 66B are bar graphs showing results for drivers with multiple domain additions in combination with improving amino acid substitutions, having amino acid sequences set forth in SEQ ID NOS: 1218 and 1220-1232.

[0093] FIGS. 67A and 67B are bar graphs showing results for drivers with multiple domain additions in combination with top of improving amino acid substitutions found previously (SEQ ID NO: 619, 1686 and 1688-1701).

[0094] FIGS. 68A and 68B are bar graphs showing results for drivers with multiple domain additions in combination with improving amino acid substitutions found previously, having amino acid sequences set forth in SEQ ID NOS: 1423-1439.

[0095] FIGS. 69A and 69B are bar graphs showing results for drivers with multiple domain additions in combination with improving amino acid substitutions having amino acid sequences set forth in any one of SEQ ID NOS: 1394-1385, 1702-1711 and 1714-1715.

[0096] FIGS. 70A and 70B are bar graphs showing results for drivers with the GFP reporter construct having a nucleic acid sequence set forth in SEQ ID NO: 175.

[0097] FIGS. 71A and 71B illustrate the results of integration assays using different R2 retrotransposable elements in U2OS cells. The different retrotransposable element constructs were used in the trans configuration (driver and GFP reporter).

[0098] FIGS. 72A and 72B are bar graphs showing results of integration assays using engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOS: 636, 1799 and 1822-1834.

[0099] FIGS. 73A and 73B are bar graphs showing results of integration assays using engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOS: 251, 620-622, 636, 1798 and 1821.

[0100] FIG. 74 is a bar graph showing results of integration assays using engineered R2-TG drivers amino acid substitution having amino acid sequences set forth in SEQ ID NOS: 629, 636, 1599, 1769, 1795-1798, 1800-1805.

[0101] FIGS. 75A and 75B are bar graphs showing results of integration assays using engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOS: 623, 624, 636, 1806-1820.

[0102] FIG. 76A is a bar graph showing results of integration assays using engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOs: 1401, 1406. FIG.

[0103] 76B is a bar graph showing results of integration assays engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOs: 1402, 1405, 1407, 1410.

[0104] FIG. 76C is a bar graph showing results of integration assays using engineered R2-TG having amino acid sequences set forth in SEQ ID NOs: 1403, 1404.

[0105] FIG. 76D is a bar graph showing results of integration assays using engineered R2-TG having amino acid sequences set forth in SEQ ID NOs: 1417-1421.

[0106] FIG. 77A is a bar graph showing results of integration assays using engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOs: 1368-1372 and 1374-1375.

[0107] FIG. 77B is a bar graph showing results of integration assays using engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOS: 139, 139, 234, 258, 1376, 1378, 1380-1389 and 1396.

[0108] FIG. 77C is a bar graph showing results of integration assays using engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOs: 1390, 1378, 1391, 1218, 636, 260, 618, 1392-1398, 1375, 1388, 1228.

[0109] FIG. 77D is a graph showing results of integration assays using engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOs: 636, 618, 783, and 1234.

[0110] FIGS. 77E and 77F are graphs showing the results of integration assays using engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOs: 636, 936, 1021, 618, 783, 1234-1236, 3028, 3029.

[0111] FIG. 78A is a graph showing results of integration assays using engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOs: 636, 618, 783, and 1234 FIG.

[0112] 78B is a graph showing results of integration assays using engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOs: 636, 618, 783, 1234.

DETAILED DESCRIPTION

[0113] The disclosure provides methods and compositions useful to promoting the integration of a transgene into a genome of a cell. Without being bound by theory or mechanism, the inventors show that by using nucleotide constructs that encode an engineered driver protein and a transgene (e.g., GFP), the transgene can be integrated into the genome of cells.

I. Engineered driver proteins

[0114] Retrotransposable elements, which are also be referred to herein interchangeably as “retroelements” or “RTEs”, or “retrotransposons” are parasitic nucleotide sequences that have naturally evolved to facilitate the incorporation of nucleotide sequences, e.g., in nature a copy of itself, into a genome of a cell, by taking advantage of the cell’s own nucleotide repair machinery to replicate themselves. A RTE typically encode a protein, which may be referred to herein as a “RTE protein”, that includes a reverse transcriptase (RT) domain and an endonuclease (EN) domain. The RT domain is responsible for synthesizing complementary DNA (cDNA) from a RNA template. In nature, the RNA template is generally a RNA transcript (which may be referred to herein as a “RTE transcript”) of the RTE, so that the reverse transcription results in replication of the original RTE in cDNA form. Typically (but without being held to theory or mechanism), the RTE transcript comprises one or more untranslated regions (RTE-UTRs) that, as RNA, forms a structure is capable of binding the RTE protein e.g., at a compatible RNA binding domain (RBD) found in the RTE protein, to facilitate the reverse transcription of the RTE transcript by the RTE protein. Typically, it is thought that the RTE protein and the one or more RTE-UTRs of a given RTE have co-evolved to have sufficient binding affinity for each other to allow them to bind to each other in the environment of a cell interior, e.g., in the nucleus. Without being held to theory or mechanism, the EN domain facilitates the insertion of the synthesized DNA into the host genome by creating a double-strand break at a target site within genomic DNA of a cell. After the double strand break, the synthesized DNA is integrated into the host cell’s genomic DNA through one or more aspects of the host cell’s own genomic repair mechanisms. However, these RT and EN mechanisms can be used to incorporate other templates with nucleotide sequences encoding a gene of interest, i.e., a transgene.

[0115] As used herein, a “driver protein” (or simply a “driver”) refers to a RTE protein, or an engineered variant thereof comprising a RTE protein or a portion thereof. In some embodiments,

the RTE protein may be optionally fused with one or more heterologous polypeptides (as described in further detail hereinbelow). The engineered driver proteins of the disclosure may retain the same or substantially the same function as the reference RTE protein, e.g., at least a RT activity and an EN activity.

[0116] As used herein, a “polypeptide” of a protein refers to either a full-length version of a protein or to a portion thereof. A polypeptide of a protein may also comprise one or more amino acid substitutions with respect to the protein.

[0117] As used herein, a “RTE polypeptide” refers to a full-length RTE protein or to a portion thereof. The portion may be, for example, a functional domain of the protein. Alternatively, the portion may be what remains of a RTE protein after deletion of one or more amino acids, segments, functional domains, binding domains, motifs, or combinations thereof. A “RTE polypeptide” as used herein may also include a full-length RTE protein or a portion thereof that has been subjected to one or more amino acid substitutions. For example, what remains of a RTE protein after deletion of the first 100 amino acids from the N-terminus, along with substitutions of certain amino acids in the RT domain and EN domain, may be referred to herein as a RTE polypeptide. In another example, what remains of a RTE protein after deletion of the first 100 amino acids from the N-terminus, as well as the RT domain, may be referred to herein as a RTE polypeptide.

[0118] As will be described in further detail herein below, the present disclosure provides embodiments of engineered drivers, comprising a RTE protein or portion thereof (e.g., a RTE polypeptide), optionally fused to one or more heterologous polypeptides, that can genomically incorporate template nucleotides with an improvement in at least one of efficiency, specificity, accuracy, fidelity or processivity compared to a reference driver, e.g., a naturally occurring, wild type (WT) RTE protein. The present disclosure also provides nucleic acids (which may be interchangeably referred to herein as “polynucleotides”, and may be DNA, RNA or a DNA/RNA hybrid) that encode the engineered drivers.

RTEs

[0119] Engineered drivers of the disclosure may comprise a RTE polypeptide encoded in one of various RTEs, including those known in the art, as well as new RTEs described herein.

[0120] In certain embodiments, the RTE is a non-LTR retroelement. In certain embodiments, the non-LTR retroelement is a long-interspersed element polypeptide (LINE) or a short-interspersed element (SINE). LINEs (Long INterspersed Elements) and SINEs (Short INterspersed Elements) are non-LTR retroelements that are found in almost all eukaryotes and are among the most common retroelements in the human genome. Wild-type LINEs typically encode a protein with reverse transcriptase and endonuclease activities. SINEs do not encode reverse transcriptase or endonuclease and depend on reverse transcriptase and endonuclease encoded by partner LINEs. In certain embodiments, the LINE is a LINE-1, a LINE-2, or a LINE-3.

[0121] In certain embodiments, the RTE polypeptide is derived from a Class I transposable element that utilizes an intermediate RNA during integration by a copy-and-paste mechanism, with or without Long Terminal Repeats (LTR) or Inverted Terminal Repeats (ITR) in their structure, and that can be derived from a superclade selected from the group consisting of retrons (include clades AbiA, AbiK, AbiP2, CRISPR, CRISPR-like, DCRs, G2L4, G2L, G2Lb, G2Lc, GII, Retrons, UG1, UG2, UG3, UG4, UG5, UG6, UG7, UG8, UG9, UG10, UG11, UG12, UG13, UG14, UG15, UG16, UG17, UG18, UG19, UG20, UG21, UG22, UG23, UG24, UG25, UG26, UG27, UG28 and UG28b), Group II introns (including clades E, ML, B, CL1A, CL1B, CL2A, CL2B, F, D and C), non-LTR retroelements (including clades CRE, R2 group (including R4 clade and R2 clade), Hero, NeSL, RandI, Proto1, L1, Tx1, RTEPT, Proto2, RTE, RTE, Outcast, Ingi (which includes sub-clade Vingi), I, Nimb, Tad1, Loa, R1, Jockey, Rex1, CR1, L2, L2A, L2B, Daphne, and Crack), LTR-retroelements or Endogenous Retroviruses (ERV) (including clades Ty1/Copia, Hepadnaviridae, BEL/Pao, Retroviridae, Caulimoviridae, and Ty3/Gypsy), Penelope-like retroelements (including clades Pen1, Pen1a, Pen2, Pen3, Pen3a, Pen3b, Pen4, Pen5, Pen5a and Pen6), Tyrosine recombinase retroelements (including clades CryptonA, CryptonF, CryptonS and CryptonI) and DIRS retroelements (including clades DIRS and Ngaro).

[0122] In exemplary embodiments, the RTE polypeptide is derived from a RTE selected from the group consisting of CRE, R2 group (including R4 clade and R2 clade), Hero, NeSL, RandI, Proto1, L1, Tx1, RTEPT, Proto2, RTE, RTE, Outcast, Ingi (which includes sub-clade Vingi), I, Nimb, Tad1, Loa, R1, Jockey, Rex1, CR1, L2, L2A, L2B, Daphne, and Crack.

R2 group RTEs

[0123] R2 group RTEs in nature are non-long terminal repeat (non-LTR) retroelements that specifically insert DNA sequences into ribosomal RNA (rRNA) gene loci in the genomic DNA of a host cell. R2 group RTEs, which include members of the R2 clade and R4 clade, typically include a single open reading frame (ORF) that encodes a single protein. As used herein, RTEs of the R2 clade as well as the R4 clade may be referred to herein as “R2 RTEs”.

[0124] Non-R2 group RTEs, including LTR RTEs and non-LTR RTEs enumerated above that are not in the R2 group, also typically encode a RTE protein that comprises a reverse transcriptase domain and an endonuclease domain, and are referred to herein as “non-R2 RTE proteins”. These non-R2 group RTEs may have, depending on their classification, one or more additional ORFs that encode chaperone proteins that have certain RNA- or DNA-binding domains. Such additional proteins are referred to herein as “non-R2 chaperone proteins”.

[0125] A RTE protein encoded in R2 group RTE’s, including R2 clade and R4 clade RTEs, may be referred to herein as a “R2 RTE protein”, or a “R2 protein”. A R2 RTE protein or a portion thereof may be referred to herein as a “R2 RTE polypeptide” or a “R2 polypeptide”.

[0126] In certain embodiments, an engineered driver comprises a R2 polypeptide, that is, a polypeptide of a R2 protein. In certain embodiments, the R2 polypeptide may be fused to one or more heterologous polypeptides, of which various embodiments are described herein.

[0127] R2 RTEs are widely distributed across various animal species, with a presence in invertebrates such as insects and mollusks, as well as in vertebrates such as birds, reptiles, and amphibians. R2 RTEs are commonly integrated into the 28S ribosomal RNA (rRNA) genes and are maintained in a stable, lineage-specific manner across generations. R2 RTE proteins from different animal species show sequence variation, reflecting the adaptation of R2 elements to different host genomes and their co-evolution. Also note that a particular species of animal may have multiple versions of R2 group RTEs, each with different nucleotide sequences (as well as different amino acid sequences for their respective RTE proteins), integrated within their genome.

[0128] Each unique RTE from a given animal species may be described herein with the following nomenclature: A clade identifier (e.g. R2 or R4) followed by the name of the

organisms from which the RTE originated, or a two- to four-letter abbreviation thereof. As such, for example, a RTE from clade R2 of *Taeniopygia guttata* (zebra finch), may be referred to herein as a “R2-TG”, and a RTE from clade R4 of *Anolis carolinensis* (green anole), may be referred to herein as a “R4-AC”. In instances where multiple RTEs have been identified in a given organism species, the clade identifier and the species abbreviation may be interspersed with a counter numeral. As such, a first RTE from clade R2 of *Chamaelys choloensis* may be referred to as R2-1-ChCh (or R2_1-ChCh), and a second RTE from clade R2 of *Chamaelys choloensis* may be referred to as R2-2-ChCh (or R2_2-ChCh). RTE proteins and polypeptides, and nucleic acids encoding them, may be referred to in a similar manner. As such, for example, a R2 protein encoded in the R2-1-ChCh RTE may be referred to herein as a R2-1-ChCh protein, and a R2 polypeptide of a R2 protein encoded in the R2-1-ChCh RTE may be referred to herein as a R2-1-ChCh polypeptide.

[0129] In certain embodiments, a R2 polypeptide of a R2 protein may be the wildtype (WT) version of the corresponding R2 protein. In certain embodiments, a R2 polypeptide of a R2 protein may be a portion of the WT version of the corresponding R2 protein. In certain embodiments, the R2 polypeptide of a R2 protein may comprise one or more amino acid substitutions compared to the WT version of the corresponding R2 protein. As used herein, a WT version of a R2 protein may be referred to as a “WT R2 protein”.

[0130] In certain embodiments, the R2 polypeptide may have least 70%, at least 71%, at least 72%, at least 73%, at least 74%, at least 75%, at least 76%, at least 77%, at least 78%, at least 79%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% sequence identity with the WT version of the corresponding R2 protein.

[0131] The percent identity between two such polypeptides (as recited herein above and elsewhere throughout to present disclosure) can be determined by manual inspection of the two optimally aligned polypeptide sequences or by using software programs or algorithms (e.g., CLUSTAL, MUSCLE, MAFFT) using standard parameters, familiar to one with skill in the art. % Sequence identity may be calculated in one of a number of ways. % Sequence identity may be calculated in one of a number of ways. In certain embodiments, the % sequence identity may be a % subject identity, which is the percentage of identical matches between a query sequence and

a subject sequence, relative to the length of the query sequence. In certain embodiments, the % sequence identity may be a % query identity, which is the percentage of identical matches between a query sequence and a subject sequence, relative to the length of the subject sequence. In certain embodiments, the % sequence identity may be a % alignment identity, which restricts the calculation of the % identity to the region(s) of alignment, so that large deletions or added domains are discounted.

[0132] In certain embodiments, an engineered driver of the disclosure may comprise a RTE polypeptide of one of the following R2 proteins, as listed in Table 1:

Table 1: R2 RTE protein sequences (amino acid and DNA sequence) of representative RTEs

	RTE name	Organism species	Amino acid sequence of RTE protein SEQ ID NO	Nucleotide sequence encoding RTE protein SEQ ID NO
1	R2-1-TG*	Taeniopygia guttata	1	10001
2	R2-ZA	Zonotrichia albicollis	2	
3	R4-AC	Anolis carolinensis	3	10003
4	R4-1-PH	Parhyale hawaiiensis	740	20752

* R2-1-TG is also referred to herein as “R2-1_TG” or “R2-TG”, which are all to be understood to refer to the same RTE, with the WT RTE protein having the amino acid sequence set forth in SEQ ID NO: 1, and encoded by the DNA sequence set forth in SEQ ID NO: 10001 (or a RNA version thereof).

[0133] Table 1 provides the amino acid and DNA sequences of an R2 protein from each of the listed RTE species. For each DNA sequence provided, it will be understood that the present disclosure also includes a respective RNA version thereof.

[0134] In certain embodiments, an engineered driver of the disclosure may comprise a RTE polypeptide of one of the following R2 proteins, as listed in Table 2:

Table 2: R2 RTE protein sequences (amino acid and DNA sequence) of R2 RTEs

	RTE name	Abbreviation	Organism species	Amino acid sequence of RTE protein SEQ ID NO	Nucleotide sequence of RTE protein SEQ ID NO
1	R2-1_Acridotheres tristis	R2-AT	Acridotheres tristis	30001	31001
2	R2-1_Acrocephalus scirpaceus	R2-ASS	Acrocephalus scirpaceus	30002	31002
3	R2-1_Agelaius tricolor	R2-AgT	Agelaius tricolor	30003	31003
4	R2-1_Ammodramus caudacutus	R2-AC	Ammodramus caudacutus	30004	31004
5	R2-1_Ammospiza nelsoni	R2-AN	Ammospiza nelsoni	30005	31005
6	R2-1_Buteo platypterus	R2-BP	Buteo platypterus	30006	31006
7	R2-1_Chamaeetus choloensis	R2-CC	Chamaeetus choloensis	30007	31007
8	R2-1_Chamaeetus fuelleborni	R2-CF	Chamaeetus fuelleborni	30008	31008
9	R2-1_Cichladusa arquata	R2-CA	Cichladusa arquata	30009	31009
10	R2-1_Ciconia episcopus	R2-CE	Ciconia episcopus	30010	31010
11	R2-1_Colaptes auratus	R2-CoAu	Colaptes auratus	30011	31011
12	R2-1_Cossypha archeri	R2-CoAr	Cossypha archeri	30012	31012
13	R2-1_Cossypha dichroa	R2-CD	Cossypha dichroa	30013	31013
14	R2-1_Falco fasciinucha	R2-FF	Falco fasciinucha	30014	31014
15	R2-1_Gallinago paraguayae	R2-GP	Gallinago paraguayae	30015	31015
16	R2-1_Garrulus glandarius	R2-GG	Garrulus glandarius	30016	31016
17	R2-1_Geranoetus albicaudatus	R2-GA	Geranoetus albicaudatus	30017	31017
18	R2-1_Harpypopsis novaeguineae	R2-HN	Harpypopsis novaeguineae	30018	31018
19	R2-1_Limnodromus scolopaceus	R2-LS	Limnodromus scolopaceus	30019	31019
20	R2-1_Melierax canorus	R2-MC	Melierax canorus	30020	31020
21	R2-1_Myiozetetes cayanensis	R2-MyC	Myiozetetes cayanensis	30021	31021
22	R2-1_Otis tarda	R2-OT	Otis tarda	30022	31022
23	R2-1_Prinia subflava	R2-PS	Prinia subflava	30023	31023

	RTE name	Abbreviation	Organism species	Amino acid sequence of RTE protein SEQ ID NO	Nucleotide sequence of RTE protein SEQ ID NO
24	R2-1_Serinus_canaria	R2-SC	Serinus canaria	30024	31024
25	R2-1_Aphelocoma_californica	R2-ApC-1	Aphelocoma californica	30025	31025
26	R2-1_Oenanthe_melanoleuca	R2-OM	Oenanthe melanoleuca	30026	31026
27	R2-1_Phyloscopus_trochilus_trochilus	R2-PTT	Phylloscopus trochilus trochilus	30027	31027
28	R2-1_Geothlypis_trichas	R2-GT	Geothlypis trichas	30028	31028
29	R2-1_Xenus_cinereus	R2-XC	Xenus cinereus	30029	31029
30	R2-2_Aphelocoma_californica	R2-2-ApC	Aphelocoma californica	30030	31030
31	R2-2_Cichladusa_arquata	R2-CiA	Cichladusa arquata	30031	31031
32	R2-1_Scolopax_minor	R2-SM	Scolopax minor	30032	31032
33	R2-1_Tyto_alba	R2-Ta	Tyto alba	30033	31033
34	R2-2_Acrocera_orbiculus	R2-AcC	Acrocera orbiculus	30034	31034
35	R2-1_Drosophila_sulfurigaster_sulfurigaster	R2-DSS	Drosophila sulfurigaster sulfurigaster	30035	31035
36	R2-1_Rhopalosiphum_padi	R2-RP	Rhopalosiphum padi	30036	31036
37	R2-1_Nemotelus_nigrinus	R2-NN	Nemotelus nigrinus	30037	31037
38	R2-1_Chlorops_oryzae	R2-CO	Chlorops oryzae	30038	31038
39	R2-2_Andrena_bucephala	R2-AB	Andrena bucephala	30039	31039
40	R2-3_Ascidia_mentula	R2-AM	Ascidia mentula	30040	31040
41	R2-1_Crotalus_oreganus_helleri	R2-COH	Crotalus oreganus helleri	30041	31041
42	R2-1_Gopherus_flavomarginatus	R2-GF	Gopherus flavomarginatus	30042	31042
43	R2-1_Malaclemys_terrapi_n_pileata	R2-MTP	Malaclemys terrapi_n pileata	30043	31043
44	R2-1_Eublepharis_macularius	R2-EM	Eublepharis macularius	30044	31044
45	R2-1_Emys_orbicularis	R2-EO	Emys orbicularis	30045	31045

	RTE name	Abbreviation	Organism species	Amino acid sequence of RTE protein SEQ ID NO	Nucleotide sequence of RTE protein SEQ ID NO
46	R2-1_Caretta_caretta	R2-CC	Caretta caretta	30046	31046
47	R2-1_Pleurodeles_waltl	R2-PW	Pleurodeles waltl	30047	31047
48	R2-1_Aspidoscelis_tigris_stejnegeri	R2-1-ATS	Aspidoscelis tigris stejnegeri	30048	31048
49	R2-1_Aspidoscelis_tigris_stejnegeri	R2-2-ATS	Aspidoscelis tigris stejnegeri	30049	31049
50	R2-1_Mauremys_mutica	R2-1-MM	Mauremys mutica	30050	31050
51	R2-3_Mauremys_mutica	R2-1-MM	Mauremys mutica	30051	31051
52	R2-1_Xiphophorus_hellerii	R2-XH	Xiphophorus hellerii	30052	31052
53	R2-1_Gambusia_affinis	R2-GamA	Gambusia affinis	30053	31053
54	R2-1_Schizopygopsis_malacanthus	R2-SMa	Schizopygopsis malacanthus	30054	31054
55	R2-1_Schizopygopsis_pylzovi	R2-1-SP	Schizopygopsis pylzovi	30055	31055
56	R2-2_Schizopygopsis_pylzovi	R2-2-SP	Schizopygopsis pylzovi	30056	31056
57	R2-1_Gila_orcuttii	R2-GO	Gila orcuttii	30057	31057
58	R2-1_Ahaetulla_prasina	R2-AP	Ahaetulla prasina	30058	31058
59	R2-1_Aldabrachelys_gigantea	R2-AG	Aldabrachelys gigantea	30059	31059
60	R2-1_Ascidia_mentula	R2-AsMe	Ascidia mentula	30060	31060

[0135] Table 2 provides the amino acid and DNA sequences of an R2 protein from each of the listed RTE species. For each DNA sequence provided, it will be understood that the present disclosure also includes a respective RNA version thereof.

[0136] In certain embodiments, an engineered driver of the disclosure may comprise a R2 polypeptide of one of the following R2 proteins: R2-XC protein (SEQ ID NO:30029), R2-LS protein (SEQ ID NO: 30019), R2-MC protein (SEQ ID NO: 30020), R2-GP protein (SEQ ID NO: 30015), or R2-PTT protein (SEQ ID NO: 30027).

[0137] In certain embodiments, an engineered driver comprises a non-R2 RTE polypeptide. In certain embodiments, the non-R2 RTE polypeptide may be fused to one or more heterologous polypeptides, of which various embodiments are described herein.

Engineered drivers

[0138] As noted above, the present disclosure provides engineered drivers, comprising a RTE protein or portion thereof, and which may be optionally fused to one or more heterologous polypeptides. These engineered driver of the disclosure can genomically incorporate template nucleotides with an improvement in at least one of efficiency, specificity, accuracy, fidelity or processivity compared to a reference driver, e.g., a naturally occurring, wild type (WT) RTE protein.

[0139] The term “efficiency” with respect to gene delivery as used herein refers to the percent of total transgene integration at a target site in the genome. Efficiency of genomic integration can be measured, for example, by amplicon sequencing and comparing the number of insertions to non-insertions at a target site and characterizing as a percentage, or by genomically integrating a transgene encoding a fluorescent protein (e.g., GFP), and measuring a percentage of fluorescent protein-positive cells. As used herein, “efficiency” and “activity” are used interchangeably when used with respect to genomic integration of transgenes by engineered drivers of the disclosure. For example, in certain working Examples provided below, embodiments of the engineered drivers of the disclosure demonstrated their “enhanced” or “improved” activity through improved efficiency in the genomic integration of transgenes encoded in gene delivery constructs. In certain embodiments, engineered drivers of the disclosure may exhibit improvement in efficiency of genomic integration by at least 1.5x fold, at least 2x fold, at least 2.5x fold, at least 3.5x fold, at least 4x fold, at least 4.5x fold, at least 5x fold, at least 10x fold, at least 15x fold, at least 20x fold, at least 25x fold, at least 30x fold, at least 40x fold, at least 50x fold, at least 60x fold, at least 70x fold, at least 80x fold, at least 90x fold, at least about 100x fold, at least about 150x fold, at least about 200x fold, at least about 300x fold, at least about 400x fold, at least about 500x fold, at least about 750x fold, or about 1000x fold compared to a reference driver.

[0140] The term “specificity” with respect to gene delivery as used herein refers to the fidelity of insertion at a specific target site. An engineered protein with high specificity would exhibit

few or no off-target insertions compared to insertions into a safe harbor site. Specificity may be measured by, for example, amplicon sequencing of known or predicted off-target sites.

[0141] The term “accuracy” with respect to gene delivery as used herein refers to the percentage of correct insertions (e.g., full sequence insertions) at the target site and can be measured by, for example, sequencing the target site and comparing correct insertions to total insertions at the site of interest.

[0142] The term “fidelity” with respect to gene delivery as used herein refers to the error rate as measured, for example, at a per nucleotide basis of the inserted sequence compared to the desired sequence.

[0143] The term “processivity” with respect to gene delivery as used herein is a measure of the length of insert that may be incorporated in its entirety. In some embodiments, large insertions may be desired. A large insertion may be a nucleic acid sequence of about 20 to about 10,000 bases or more. For example, about 20 bases, about 50 bases, about 100 bases, about 200 bases, about 300 bases, about 400 bases, about 500 bases, about 750 bases, about 1000 bases, about 1250 bases, about 1500 bases, about 2,000 bases, about 3,000 bases, about 4,000 bases, about 5,000 bases, about 6,000 bases, about 7,000 bases, about 8,000 bases, about 9,000 bases, about 10,000 bases or more.

Heterologous polypeptides

[0144] The present disclosure provides embodiments of engineered drivers that comprises a RTE polypeptide fused to one or more heterologous polypeptides (e.g. fused to 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more heterologous polypeptides), and also provides as nucleic acids that encode said engineered drivers fused to said heterologous polypeptides. As such, any disclosures herein regarding embodiments of engineered drivers, RTE polypeptides, and heterologous polypeptides fused thereto, should be understood as also disclosing nucleic acids that encode the embodiments of engineered drivers, RTE polypeptides, and heterologous polypeptides fused thereto.

[0145] It is to be noted that a “heterologous” polypeptide of the engineered driver refers to a polypeptide that is not a portion of the RTE protein from which the RTE polypeptide is derived, and is separate and independent from the heterologous nucleic acid (e.g., the transgene) of gene

delivery constructs of the disclosure (which is described further herein below). Furthermore, some of the polypeptides disclosed herein may be known to have more than one function and hence be present in one or more functional category. For example, Rad 51 and RAD17 are listed as both RNA/DNA processing polypeptides and RNA/DNA repair polypeptides.

[0146] In certain embodiments, the driver comprises a RTE polypeptide fused to at least one heterologous polypeptide at the N-terminus of the RTE polypeptide. In certain embodiments, the driver comprises a RTE polypeptide fused to at least one heterologous polypeptide at the C-terminus of the RTE polypeptide. In certain embodiments, the driver comprises a RTE polypeptide fused to at least one heterologous polypeptide internally within the RTE polypeptide. In certain embodiments, the driver comprises a RTE polypeptide fused to at least one heterologous polypeptide, independently, at the N-terminus of the RTE polypeptide and at the C-terminus of the RTE polypeptide. In certain embodiments, the driver comprises a RTE polypeptide fused to at least one heterologous polypeptide internally within the RTE polypeptide and to at least one heterologous polypeptide, independently, at the N-terminus and or the C-terminus of the RTE polypeptide.

[0147] In certain embodiments, the at least one heterologous polypeptide comprises a heterologous nucleosome binding polypeptide, a heterologous nucleic acid binding polypeptide, a heterologous RNA/DNA repair polypeptide, a heterologous RNA/DNA processing polypeptide, heterologous localization signal, a heterologous reverse transcriptase, a heterologous endonuclease, a heterologous intrinsically disordered polypeptide, or any combination thereof.

[0148] In certain embodiments, in an engineered driver comprising a plurality of heterologous polypeptides in sequence, pairs of adjacent heterologous polypeptides may be connected directly, or via a linker. In certain embodiments, a heterologous polypeptide and a RTE polypeptide situated adjacent to each other may be directly connected, or connected via a linker. In certain embodiments, the linker is a rigid linker. In other embodiments, the linker is a flexible linker. Flexible linkers are generally made up of small, non-polar (e.g., Gly) or polar (e.g., Ser or Thr) amino acids. Alternating Gly and Ser residues provides flexibility. In certain embodiments, the linker is a glycine-serine based linker or a XTEN peptide linker. In certain embodiments, the linker is a cleavable linker. Solubility of the linker and associated sequences may be enhanced by the inclusion of charged residues, e.g., two positively charged residues (e.g., Lys) and one

negatively charged residue (e.g., Glu). The linker may be 2-35 amino acids long. In certain embodiments, the linker is selected from any one of SEQ ID NOs: 102-104.

[0149] In certain embodiments, the at least one heterologous polypeptide comprises an amino acid sequence set forth in any one of SEQ ID NOS: 21-61, 295, 301, 583-588, 785-789, and 1171-1216, or an amino acid sequence that is at least 70% identical, at least 75% identical, at least 80% identical, at least 85% identical, at least 90% identical, at least 91% identical, at least 92% identical, at least 93% identical, at least 94% identical, at least at least 95% identical, at least 96% identical, at least 97% identical, at least 98% identical, or 99% identical, to a sequence of any of SEQ ID NOs: 21-61, 295, 301, 583-588, 785-789, and 1171-1216.

Arrangement of heterologous polypeptides fused to a RTE polypeptide

[0150] Non-limiting examples of engineered drivers that can be used (e.g., directly or encoded on an RNA and/or DNA molecule) to promote insertion of a heterologous gene into a target nucleic acid in a host cell include protein fusions comprising a RTE polypeptide, optionally fused to at least one heterologous polypeptide that redirects and/or enhances insertion of the heterologous gene. In some embodiments, each of the at least one heterologous polypeptide comprises one or more of, e.g., an RNA/DNA processing polypeptide, an RNA/DNA repair polypeptide, a nucleic acid binding polypeptide, or a nucleosome binding polypeptide. The at least one heterologous polypeptide can be fused to the N-terminus and/or C-terminus of a RTE polypeptide, and/or internally within a RTE polypeptide (e.g., between two domains of the RTE polypeptide). Nucleic acids (e.g., RNA and/or DNA) encoding one or more engineered drivers can be used to promote insertion of a transgene (e.g., a heterologous nucleic acid comprising a gene of interest) flanked by terminal regions (e.g., by 5' and 3' terminal regions comprising, e.g., RTE-UTRs, homology arms, etc.) into a target nucleic acid. In some embodiments, a first nucleic acid (a gene delivery construct) that includes a heterologous nucleic acid (e.g., heterologous gene) flanked by RTE-UTRs, along with a second nucleic acid (driver construct) that encodes an engineered driver is provided (e.g., delivered/administered to a cell or subject). In some embodiments, the nucleic acid that encodes the gene delivery construct and the driver construct are provided (e.g., administered to a subject) as separate nucleic acid molecules (e.g., in trans), or on a single nucleic acid molecule (i.e., in cis).

[0151] The following schemes represent non-limiting exemplary configurations of one or more heterologous polypeptides (HPs) fused to RTE (e.g., R2 RTE) polypeptides (e.g., to RTE domains, e.g., an endonuclease domain (EN) and/or a reverse transcriptase domain (RT). In the below schemes, an N-terminally fused heterologous polypeptide is referred to as an nHP. A C-terminally fused heterologous polypeptide is referred to as a cHP. An internally fused heterologous polypeptide is referred to as an iHP. N- indicates the N-terminus of the fusion protein, and -C indicates the C-terminus of the fusion protein. In some embodiments, a RTE polypeptide is modified to remove the cognate endonuclease domain (e.g., to remove a restriction-like endonuclease (RLE) domain). In some embodiments, a heterologous endonuclease domain (hEN) is fused to the RT domain (e.g., to replace the cognate endonuclease domain). In some embodiments, a heterologous polypeptide is fused to an RTE polypeptide via one or more linkers. For example, a linker may be located at the N- and/or C-terminus of a heterologous polypeptide to connect it to a RTE polypeptide (e.g. R2 polypeptide) and/or to other heterologous polypeptides.

Exemplary HP fusions to RTE domains:

- i) N-nHP-RT-EN-C
- ii) N-RT-EN-cHP-C
- iii) N-nHP-RT-EN-cHP-C
- iv) N-nHP-RT-iHP-EN-C
- v) N-RT-iHP-EN-cHP-C
- vi) N-nHP-RT-iHP-EN-cHP-C

Exemplary HP fusions to a RTE comprising a heterologous endonuclease domain (hEN):

- i) N-nHP-RT-hEN-C
- ii) N-RT-hEN-cHP-C
- iii) N-nHP-RT-hEN-cHP-C
- iv) N-nHP-RT-iHP-hEN-C
- v) N-RT-iHP-hEN-cHP-C
- vi) N-nHP-RT-iHP-hEN-cHP-C

Exemplary HP fusions to RTE Domains:

- i) N-nHP-RT-C
- ii) N-RT-cHP-C
- iii) N-nHP-EN-C
- iv) N-RT-cHP-C

[0152] One or more nucleic acids (e.g., RNA and/or DNA) encoding at least one of these engineered drivers can be provided, in trans or in cis, to target cells (e.g., ex vivo or in vivo) along with one or more nucleic acids (e.g., RNA and/or DNA) encoding a transgene (e.g., flanked by terminal regions) to promote integration of the transgene into a nucleic acid (e.g., a genomic nucleic acid) of the target cells.

[0153] Non-limiting examples of different engineered driver configurations are schematically illustrated in FIGS. 2A-2L.

[0154] FIGS. 2A-2L illustrate non-limiting examples of different configurations of an engineered driver comprising a RTE polypeptide and one or more heterologous polypeptides that can promote and/or redirect integration of a transgene into a target nucleic acid (e.g., into the genome of a cell). In certain embodiments, an engineered driver comprises at least one heterologous polypeptide (e.g., comprising one or more RNA/DNA processing polypeptides, RNA/DNA repair polypeptides, nucleic acid binding polypeptides, and/or nucleosome binding polypeptides) fused to the N-terminus of a RTE polypeptide, the C-terminus of a RTE polypeptide, and/or internally (e.g., between two domains of a RTE polypeptide).

[0155] FIG. 2A illustrates a non-limiting example of an engineered driver comprising a RTE polypeptide that has a heterologous polypeptide fused to its N-terminus.

[0156] FIG. 2B. illustrates a non-limiting example of an engineered driver comprising a RTE polypeptide that has a heterologous polypeptide fused to its C-terminus.

[0157] FIG. 2C illustrates a non-limiting example of an engineered driver comprising a RTE polypeptide that has a first heterologous polypeptide fused to its N-terminus and a second heterologous polypeptide fused to its C-terminus.

[0158] In certain embodiments, each heterologous polypeptide can itself independently comprise one or more (e.g., two, three, four, or more) different polypeptides (for example one or more of an RNA/DNA processing polypeptide, RNA/DNA repair polypeptide, nucleic acid binding polypeptide, and/or nucleosome binding polypeptide), optionally along with one or more linkers and/or localization sequences.

[0159] In certain embodiments, an engineered driver may comprise a plurality of heterologous polypeptides arranged sequentially. Schematic illustrations of sequential sets of heterologous polypeptides in different arrangements are shown in FIGS. 2D-2L. As described hereinbelow with respect to FIGS. 2D-2L, each heterologous polypeptide in the sequence of heterologous polypeptides as show in FIGS. 2D-2L may individually be referred to herein as a “domain” of the sequential set of heterologous polypeptides. For example, each of “domain N1”, “domain N2”, “domain C1”, “domain C2” may be, individually, a heterologous polypeptide of the disclosure.

[0160] FIG. 2D schematically illustrates a non-limiting embodiment of an engineered driver comprising a RTE polypeptide fused to a sequential set of N-terminal heterologous polypeptides fused to the RTE polypeptide, with the sequential set of N-terminal heterologous polypeptides comprising a first domain and a second domain (domain N1 and domain N2).

[0161] FIG. 2E schematically illustrates a non-limiting embodiment of an engineered driver comprising a RTE polypeptide fused to a sequential set of N-terminal heterologous polypeptides comprising, in a N-terminal to C-terminal direction, a nuclear localization sequence (NLS), a domain N1, and a linker connecting the domain N1 and the RTE polypeptide.

[0162] FIG. 2F schematically illustrates a non-limiting embodiment of an engineered driver comprising a RTE polypeptide and a sequential set of C-terminal heterologous polypeptides comprising, in a N-terminal to C-terminal direction, a linker, a domain C1, and an NLS.

[0163] FIG. 2G schematically illustrates a non-limiting example of an engineered driver comprising a sequential set of N-terminal heterologous polypeptides fused to a RTE polypeptide on the N-terminal side, as well as a sequential set of C-terminal heterologous polypeptides fused to the RTE polypeptide on the C-terminal side. The sequential set of N-terminal heterologous polypeptides comprises, in a N-terminal to C-terminal direction, an optional nuclear localization

sequence (NLS), a domain N1, and a linker connecting the domain N1 and the RTE polypeptide. The sequential set of C-terminal heterologous polypeptides comprises, in a N-terminal to C-terminal direction, a linker, a domain C1, and an optional NLS.

[0164] FIG. 2H schematically illustrates a non-limiting embodiment of an engineered driver comprising a RTE polypeptide fused to an extended sequential set of N-terminal heterologous polypeptides fused to the RTE polypeptide. The extended sequence set of N-terminal heterologous polypeptides comprises, in an N-terminal to C-terminal direction, an optional NLS, a first domain N1 and a second domain N2, with a first optional linker connecting domains N1 and N2, and second optional linker connecting domain N2 to the N-terminal side of the RTE polypeptide.

[0165] FIG. 2I schematically illustrates a non-limiting embodiment of an engineered driver comprising a RTE polypeptide fused to an extended sequential set of C-terminal heterologous polypeptides fused to the RTE polypeptide. The extended sequence set of C-terminal heterologous polypeptides comprises, in an N-terminal to C-terminal direction, a first domain C1, a second domain C2 and an optional NLS, with a first optional linker connecting domain C1 to the C-terminal side of the RTE polypeptide, and second optional linker connecting domains C1 and C2.

[0166] FIG. 2J schematically illustrates a non-limiting example of an engineered driver comprising an extended sequential set of N-terminal heterologous polypeptides fused to a RTE polypeptide on the N-terminal side, as well as an extended sequential set of C-terminal heterologous polypeptides fused to the RTE polypeptide on the C-terminal side.

[0167] FIG. 2K schematically illustrates a non-limiting example of an engineered driver comprising an extended sequential set of N-terminal heterologous polypeptides (two domains N1 and N2) fused to a RTE polypeptide on the N-terminal side, as well as a shorter sequential set of C-terminal heterologous polypeptides with a single domain C1 fused to the RTE polypeptide on the C-terminal side.

[0168] FIG. 2L schematically illustrates a non-limiting example of an engineered driver comprising an extended sequential set of C-terminal heterologous polypeptides comprising two domains C1 and C2 fused to a RTE polypeptide on the C-terminal side, as well as a shorter

sequential set of N-terminal heterologous polypeptides with a single domain N1 fused to the RTE polypeptide on the C-terminal side.

[0169] It will be appreciated that other configurations can include additional and/or different elements that are fused to the N-terminus of a RTE polypeptide, the C-terminus of the RTE polypeptide, and/or internally with the RTE polypeptide.

[0170] Various embodiments of heterologous polypeptides that may be fused to a RTE polypeptide in an engineered driver of the disclosure are described in more detail hereinbelow.

Heterologous nucleosome binding polypeptide

[0171] Nucleosome binding polypeptides are characterized by being capable of binding a nucleosome, a section of genomic DNA that is wrapped around a core of histone proteins, which forms a unit of DNA packaging in eukaryotes. Without being bound by theory, nucleosome binding polypeptides have various functions that affect access to genomic DNA, including modulating chromatin structure and chromatin accessibility, and altering the activity of genomic editing proteins

[0172] In certain embodiments, the at least one heterologous polypeptide fused to a RTE polypeptide comprises a nucleosome binding polypeptide. In certain embodiments, the nucleosome binding polypeptide may be a HMGN1 (high mobility group nucleosome-binding domain-containing protein 1) polypeptide, an HMGB1 (high mobility group protein B1) polypeptide, or a StkC (sticky-C) DNA binding domain polypeptide.

[0173] In certain embodiments, HMGN1 polypeptide may comprise an amino acid sequence set forth in SEQ ID NO: 40 or an amino acid sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 97%, at least 98%, or at least 99% identical thereto. In certain embodiments, the HMGN1 polypeptide comprises at least one amino acid substitution, wherein the at least one amino acid substitution is at the D100 residue, e.g.: D100E and/or D100G relative to SEQ ID NO: 40.

[0174] In certain embodiments, HMGB1 polypeptide may comprise an amino acid sequence set forth in SEQ ID NO: 41 or an amino acid sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 97%, at least

98%, or at least 99% identical thereto. The StkC binding domain polypeptide may comprise an amino acid sequence set forth in SEQ ID NO: 57 or an amino acid sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 97%, at least 98%, or at least 99% identical thereto.

[0175] It will be appreciated that an engineered driver of the disclosure may comprise one or more, for example 1, 2, 3, 4, or 5, heterologous nucleosome binding polypeptides, optionally in combination with one or more other heterologous polypeptides as described herein.

[0176] In certain embodiments, the engineered driver protein comprises a R2 RTE polypeptide fused to one or more heterologous nucleosome binding polypeptides, e.g., a HMGN1 polypeptide, a HMGB1 polypeptide, a StkC DNA binding domain polypeptide, or combinations thereof.

[0177] In certain embodiments, the engineered driver protein comprises a R2 RTE polypeptide fused to a HMGN1 polypeptide. The HMGN1 polypeptide may be situated internally in the R2 RTE protein, on the N-terminal side of the R2 RTE polypeptide or the C-terminal side of the R2 RTE polypeptide. SEQ ID NOS: 636, 1234-1236 set forth amino acid sequences of exemplary driver proteins comprising a HMGN1 polypeptide.

Heterologous nucleic acid binding polypeptide

[0178] In certain embodiments, the at least one heterologous polypeptide fused to a RTE polypeptide comprises a nucleic acid binding polypeptide that binds to RNA and/or DNA. In certain embodiments, a nucleic acid binding polypeptide binds to RNA. In certain embodiments, a nucleic acid binding polypeptide binds to DNA.

[0179] In certain embodiments, the nucleic acid binding polypeptide may be a non-sequence specific DNA binding polypeptide. In certain embodiments, the non-sequence specific DNA binding domain may be a Sto7d DNA binding domain polypeptide or a Sso7d DNA binding domain polypeptide. Sso7D is from thermophilic archaea *Sulfolobus solfataricus*, and the Sto7d protein, is a Sso7D counterpart from *Sulfolobus tokodaii*. In certain embodiments, the non-sequence specific DNA binding domain may be a UL42 DNA binding domain polypeptide or a Guinea pig cytomegalovirus (CMV) DNA binding domain. Without being limited by theory, non-sequence specific DNA binding domains, e.g., from the Sso7d homology group can

improve processivity of DNA polymerases and reverse transcriptases, and can be used to promote integration of a transgene by an engineered driver.

[0180] In certain embodiments, the Sto7d DNA binding domain polypeptide comprises an amino acid sequence set forth in SEQ ID NO: 42, or an amino acid sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 97%, at least 98%, or at least 99% identical thereto.

[0181] In certain embodiments, the Sso7d DNA binding domain polypeptide comprises an amino acid sequence set forth in SEQ ID NO: 23, or an amino acid sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 97%, at least 98%, or at least 99% identical thereto.

[0182] In certain embodiments, the UL42 DNA binding domain polypeptide (binding domain from herpes simplex virus 1 (HSV1)) comprises an amino acid sequence set forth in SEQ ID NO: 587, or an amino acid sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 97%, at least 98%, or at least 99% identical thereto.

[0183] In certain embodiments, the Guinea pig CMV DNA binding domain polypeptide comprises an amino acid sequence set forth in SEQ ID NO: 785, or an amino acid sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 97%, at least 98%, or at least 99% identical thereto.

[0184] In certain embodiments, the Rat CMV DNA binding domain polypeptide comprises an amino acid sequence set forth in SEQ ID NO: 786, or an amino acid sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 97%, at least 98%, or at least 99% identical thereto.

[0185] In certain embodiments, the non-sequence specific DNA binding domain may be a single-stranded (ss) DNA binding polypeptide. In certain embodiments, the ss DNA binding polypeptide may be a RPA3 DNA binding domain polypeptide, which may comprise an amino acid sequence set forth in SEQ ID NO:26, or an amino acid sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 97%, at least 98%, or at least 99% identical thereto.

[0186] In certain embodiments, the ss DNA binding polypeptide may be a RecT polypeptide, which may comprise an amino acid sequence set forth in SEQ ID NO: 38 (csRecT) or SEQ ID NO: 39 (paRecT), or an amino acid sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 97%, at least 98%, or at least 99% identical thereto.

[0187] In certain embodiments, the ss DNA binding polypeptide may be a RecA polypeptide, which may comprise an amino acid sequence set forth in SEQ ID NO: 58, or an amino acid sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 97%, at least 98%, or at least 99% identical thereto.

[0188] In certain embodiments, the ss DNA binding polypeptide may be a SSB polypeptide, which may comprise an amino acid sequence set forth in SEQ ID NO: 59, or an amino acid sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 97%, at least 98%, or at least 99% identical thereto.

[0189] In certain embodiments, the nucleic acid binding polypeptide may be a sequence specific DNA binding polypeptide. Without being bound by theory, site-specific DNA binding domains may improve site-specific integration of a transgene of the disclosure into the genome (e.g., by improving local interaction of the retroelement protein at the genomic site of interest) and can be used to introduce site-specificity into otherwise non-site-specific drivers and/or improve site-specificity for site-specific drivers.

[0190] In certain embodiments, the sequence specific DNA binding polypeptide may be a Cas nuclease. Non-limiting examples of Cas nuclease include Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas6, Cas7, Cas8, Cas9 (also known as Csn1 and Csx12), Cas10, Csy1, Csy2, Csy3, Cse1, Cse2, Csc1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx15, Csf1, Csf2, Csf3, Csf4, homologues thereof, or modified versions thereof. The Cas nuclease may be a dead Cas nuclease (e.g., a dead SpCas9 having D10A and/or H840A amino acid substitutions). In certain embodiments, sequence specific DNA binding polypeptide may be a Zinc finger DNA binding domain (e.g., a Zinc finger DNA binding domain targeting AAVS1, a transcription activator-like effector (TALE) DNA binding domain, or a dead SpuFz1 (e.g., dSpuFz1, a SpuFz1 with a D606A amino acid substitution).

[0191] In certain embodiments, the nucleic acid binding polypeptide provides sequence specificity, and may comprise a sequence set forth in any one of SEQ ID NOS: 43-54, 60, 80, 81, and 82, or an amino acid sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 97%, at least 98%, or at least 99% identical thereto.

[0192] In certain embodiments, a sequence-specific endonuclease may be used, e.g., to enhance or replace a native endonuclease domain of a RTE polypeptide. Such domains may retarget integration to a site of interest. Non-limiting examples of an endonuclease fusion/replacement include a site-specific homing endonuclease targeting the AAVS1 gene fused to RTE polypeptide deficient in endonuclease activity, either through an inactivating mutation in the endonuclease domain (e.g., by a D237A, H238 substitution, and/or D216A in an L2-2 domain or a corresponding mutation in an alternative domain), or through a deletion of the endonuclease domain. Similarly, a homing endonuclease nickase variant may be used to introduce a single-strand break instead of a double-strand break. Other non-limiting examples include a Cas nuclease or nickase fused to an endonuclease-deficient RTE polypeptide.

[0193] It will be appreciated that an engineered driver of the disclosure may comprise one or more, for example 1, 2, 3, 4, or 5, heterologous nucleic acid binding polypeptides, optionally in combination with one or more other heterologous polypeptides as described herein.

[0194] In certain embodiments, the engineered driver protein comprises a R2 RTE polypeptide fused to one or more heterologous nucleic acid binding polypeptides. In certain embodiments, the engineered driver protein comprises a R2 RTE polypeptide fused to a Sto7d DNA binding domain polypeptide or the UL42 DNA binding domain polypeptide. The fused polypeptide may be situated internally in the R2 RTE protein, on the N-terminal side of the R2 RTE polypeptide or the C-terminal side of the R2 RTE polypeptide or internal to the RTE polypeptide. SEQ ID NOS: 251, 260, 618, 1234, 675, 782 set forth amino acid sequences of exemplary driver proteins comprising a Sto7d DNA binding domain polypeptide or a UL42 DNA binding domain polypeptide.

Heterologous RNA/DNA repair polypeptide

[0195] In certain embodiments, the at least one heterologous polypeptide fused to a RTE polypeptide comprises a RNA/DNA repair polypeptide, a protein that is or interacts with a host repair protein that acts on RNA and/or DNA. In some embodiments, the DNA repair polypeptide is selected from a group consisting of a non-homologous end joining (NHEJ) pathway protein, a mismatch repair (MMR) protein, a microhomology-mediated end-joining (MMEJ) protein, a homology directed repair (HDR) pathway protein, and a DNA damage response protein.

[0196] Activation or inhibition of certain host cell proteins via a heterologous polypeptide that is fused to a RTE polypeptide (e.g., a reverse transcriptase and/or endonuclease domain) may improve transgene insertion efficiency. The precise class of host cell proteins will depend on the mechanism of integration (e.g., depending on the RTE polypeptide that is used along with the terminal regions that flank the transgene). In a non-limiting example, retroelements that integrate into a precise location in the genome may rely on an HDR-based integration mechanism, while retroelements that integrate into a random location in the genome may rely on NHEJ or MMEJ-based mechanisms. Activating the corresponding repair pathway while suppressing the alternative repair pathways may increase the efficiency of desired transgene integration.

[0197] In certain embodiments, the RNA/DNA repair polypeptide may be a RAD51 polypeptide, a RAD17 polypeptide, a BRCA2 polypeptide, a ANKRD28 polypeptide, an HSV-1 alkaline nuclease polypeptide (e.g., UL12), a proliferating cell nuclear antigen (PCNA) polypeptide, a PCNA interacting protein (PIP) motif, a MDC1 polypeptide, a MSH4 polypeptide, a SCML1 polypeptide, a CDKN2A polypeptide, a p53 inhibitor peptide, or a CTIP (CtBP-interacting protein, also known as DNA endonuclease RBBP8 or Retinoblastoma-binding protein 8) polypeptide. In certain embodiments, the RNA/DNA repair polypeptide is a CTIP polypeptide.

[0198] The MRN complex is a protein complex consisting of Mre11, Rad50 and Nbs1 (also known as Nibrin in humans) involved in detection and repair of DNA damage. Without being bound by theory, an HSV-1 alkaline nuclease polypeptide (e.g., a UL12 polypeptide) is capable of recruiting the MRN complex and may, when comprised in an engineered driver, increase transgene integration by promoting HDR and MMEJ. Without being bound by theory, a RAD17

polypeptide is capable of modulating RAD51 and promoting HDR, thereby improving site-specific transgene integration when fused to a RTE polypeptide. Without being bound by theory, a BRCA2 polypeptide is capable of modulating RAD51 and promoting HDR, thereby improving site-specific transgene integration when fused to a RTE polypeptide. Without being bound by theory, ANKRD28 polypeptide (involved in transport to the Golgi and subsequent modification and Vesicle-mediated transport), may promote HDR and may, when comprised in an engineered driver, increase transgene integration. Without being bound by theory, a MDC1 (mediator of DNA damage) polypeptide, a MSH4 (the MSH4 complex binds and stabilizes double holiday junctions) polypeptide, a SCML1 polypeptide (predicted to enable chromatin binding activity and histone binding activity), or a CDKN2A (CDK4 kinase inhibitors) polypeptide may promote HDR and therefore may, when comprised in an engineered driver, increase transgene integration.

[0199] Cellular DNA damage response pathways may act to broadly inhibit genome editing. For example p53 can cause senescence of edited cells. Without being bound by theory, inhibition of one of the pathways and/or factors within these pathways, e.g., through incorporating a p53 inhibitor/repressor polypeptide in an engineered driver, may increase transgene integration. MDM2 interacts with and represses (e.g., inhibits) p53. Without being bound by theory, a MDM2 polypeptide may enhance general transgene integration through local repression of p53 when fused to a RTE polypeptide. Without being bound by theory, a synthetic polypeptide (e.g., a peptide 14-derived peptide) that inhibits p53 can be used to enhance general transgene integration through local repression of p53 when fused to a RTE polypeptide.

[0200] In certain embodiments, the RAD17 polypeptide comprises an amino acid sequence set forth in SEQ ID NO: 34, or an amino acid sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 97%, at least 98%, or at least 99% identical thereto.

[0201] In certain embodiments, the ANKRD28 polypeptide comprises an amino acid sequence set forth in SEQ ID NO: 31, or an amino acid sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 97%, at least 98%, or at least 99% identical thereto.

[0202] In certain embodiments, the HSV-1 alkaline nuclease polypeptide comprises an amino acid sequence set forth in SEQ ID NO: 22, or an amino acid sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 97%, at least 98%, or at least 99% identical thereto.

[0203] In certain embodiments, the BRCA2 polypeptide comprises an amino acid sequence that set forth in SEQ ID NO: 55, or an amino acid sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 97%, at least 98%, or at least 99% identical thereto.

[0204] In certain embodiments, the PCNA polypeptide comprises an amino acid sequence that is set forth in SEQ ID NO: 36 or SEQ ID NO: 37, or an amino acid sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 97%, at least 98%, or at least 99% identical thereto.

[0205] In certain embodiments, the MDC1 polypeptide comprises an amino acid sequence that set forth in SEQ ID NO: 32, or an amino acid sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 97%, at least 98%, or at least 99% identical thereto.

[0206] In certain embodiments, the MSH4 polypeptide comprises an amino acid sequence set forth in SEQ ID NO: 33, or an amino acid sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 97%, at least 98%, or at least 99% identical thereto.

[0207] In certain embodiments, the SCML1 polypeptide comprises an amino acid sequence set forth in SEQ ID NO: 35, or an amino acid sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 97%, at least 98%, or at least 99% identical thereto.

[0208] In certain embodiments, the CDKN2A polypeptide comprises an amino acid sequence set forth in SEQ ID NO: 36, or an amino acid sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 97%, at least 98%, or at least 99% identical thereto.

[0209] In certain embodiments, the p53 inhibitor is a MDM2 polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 28, or an amino acid sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 97%, at least 98%, or at least 99% identical thereto.

[0210] In certain embodiments, the p53 inhibitor is a peptide 14-derived peptide comprising an amino acid sequence set forth in SEQ ID NO: 56, or an amino acid sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 97%, at least 98%, or at least 99% identical thereto.

[0211] In certain embodiments, the CTIP polypeptide comprises an amino acid sequence set forth in SEQ ID NO: 61, or an amino acid sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 97%, at least 98%, or at least 99% identical thereto. In certain embodiments, the CTIP polypeptide comprises at least one amino acid substitution, wherein the at least one amino acid substitution is at one or more of (e.g. at least 1,2,3,4,5,etc.of) T15I, R123K, P181L, M202L, M202F, M202T, A237L, A237I, A237N, A237K, N288D, D18F, I49C, S17D, D18C, Q46T, I2A, C198S, E155S, L281G, I186S, Q46K, or a combination thereof relative to SEQ ID NO: 61. In exemplary embodiments, the CTIP polypeptide comprises the amino acid substitutions of T15I, R123K, P181L, M202L, M202F, M202T, A237L, A237I, A237N, A237K, N288D, D18F, I49C, S17D, D18C, Q46T, I2A, C198S, E155S, L281G, I186S, and Q46K, relative to SEQ ID NO: 61.

[0212] Exemplary embodiments of engineered drivers with a mutated CTIP polypeptide with one or more of the above-noted amino acid substitutions (relative to SEQ ID NO: 61) are set forth in SEQ ID NOS: 737-745, 1063-1073.

[0213] In certain embodiments, the RNA/DNA repair polypeptide may be a proliferating cell nuclear antigen (PCNA) interaction protein (PIP) motif domain. In certain embodiments, a RTE polypeptide has a modification that inactivates its activity (e.g., inactivation of the PCNA interaction protein (PIP) motif domain of a RTE). In certain embodiments, a RTE polypeptide has a modification that activates a PCNA interacting motif (e.g., activation of the PCNA interaction protein (PIP) motif domain of a RTE). The PCNA interaction motif is believed to recruit the cellular PCNA protein which may play a role in DNA repair. In certain embodiments, the native RTE PIP motif is replaced with a PIP motif from another protein (such as p21, FEN1

or CHAF1A) which may improve PCNA recruitment. In certain embodiments, the modification comprises deletion. In certain embodiments, the modification comprises one or more mutations (e.g., point mutations). In certain embodiments, the modification is in the endonuclease domain. In certain embodiments, the modification is in the reverse transcriptase domain. In certain embodiments, the modification is in the RNA binding domain. In certain embodiments, PIP motif polypeptides (such as a p21 polypeptide, a FEN1 polypeptide or a CHAF1A polypeptide) may be added as a heterologous polypeptide in the C-terminal or N-terminal of the RTE protein, or internally in the RTE protein.

[0214] In certain embodiments, the PIP motif polypeptide is a p21 polypeptide comprising an amino acid sequence that set forth in SEQ ID NO: 56, or an amino acid sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 97%, at least 98%, or at least 99% identical thereto.

[0215] In certain embodiments, the PIP motif polypeptide is a FEN1 polypeptide comprising an amino acid sequence that set forth in SEQ ID NO: 24, or an amino acid sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 97%, at least 98%, or at least 99% identical thereto.

[0216] In certain embodiments, the PIP motif polypeptide is a CHAF1A polypeptide comprising an amino acid sequence that set forth in SEQ ID NO: 37, or an amino acid sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 97%, at least 98%, or at least 99% identical thereto.

[0217] It will be appreciated that an engineered driver of the disclosure may comprise one or more, for example 1, 2, 3, 4, or 5, heterologous RNA/DNA repair polypeptides, optionally in combination with one or more other heterologous polypeptides as described herein.

[0218] In certain embodiments, the engineered driver protein comprises a R2 RTE polypeptide fused to one or more heterologous RNA/DNA binding polypeptides. In certain embodiments, the engineered driver protein comprises a R2 RTE polypeptide fused to a CTIP binding domain polypeptide. The fused CTIP polypeptide may be situated internally in the R2 RTE protein, on the N-terminal side of the R2 RTE polypeptide or the C-terminal side of the R2 RTE

polypeptide. SEQ ID NOS: 636, and 639-667 set forth amino acid sequences of exemplary driver proteins comprising a CTIP polypeptide.

Heterologous RNA/DNA processing polypeptide

[0219] An RNA/DNA processing polypeptide is an enzyme that causes chemical changes to RNA and/or DNA, for example by promoting or retarding RNA degradation. In addition to interacting with host cell repair and DNA-damage response proteins, proteins that directly process and/or repair RNA/DNA intermediates involved in transgene integration also may, without being bound by theory, improve transgene integration efficiency. Also without being bound by theory, an RNA/DNA processing polypeptide may improve transgene integration efficiency and/or redirect transgene integration to a different target location (e.g., within the genome of a cell).

[0220] In certain embodiments, the at least one heterologous polypeptide fused to a RTE polypeptide comprises an RNA/DNA processing polypeptide. In certain embodiments, the at least one heterologous polypeptide comprises an RNA/DNA processing polypeptide selected from a RNA helicase polypeptide, a RNA helicase recruitment motif polypeptide, a RAD51 polypeptide, a RAD17 polypeptide, or a RAD6 polypeptide.

[0221] RNA helicases are enzymes that unwind RNA structures, and play a role in various cellular processes, including but not limited to transcription, RNA splicing, and translation, as well as in the regulation of gene expression. A RNA helicase recruitment motif is a specific sequence or structural feature in proteins that facilitates the binding and recruitment of RNA helicases to their target sites.

[0222] In certain embodiments, the at least one heterologous RNA/DNA processing polypeptide comprises a RNA helicase polypeptide. In certain embodiments, the RNA helicase polypeptide is a EIF4A (Eukaryotic initiation factor 4A) polypeptide. In certain embodiments, the EIF4A polypeptide comprises an amino acid set forth in SEQ ID NO: 585, or an amino acid sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 97%, at least 98%, or at least 99% identical thereto. In certain embodiments, the RNA helicase polypeptide is a Dengue virus RNA helicase polypeptide. In certain embodiments, the Dengue virus RNA helicase polypeptide comprises an amino acid set

forth in SEQ ID NO: 1875, or an amino acid sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 97%, at least 98%, or at least 99% identical thereto.

[0223] In certain embodiments, the at least one heterologous RNA/DNA processing polypeptide comprises a RNA helicase recruitment motif polypeptide. In certain embodiments, the RNA helicase recruitment motif polypeptide is a TUDOR domain polypeptide from TDRD3 which may recruit RNA helicase DHX9 (Dexh box helicase 9). In certain embodiments, the TUDOR domain polypeptide comprises an amino acid set forth in SEQ ID NO: 584, or an amino acid sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 97%, at least 98%, or at least 99% identical thereto. In certain embodiments, the RNA helicase recruitment peptide is derived from EIF4G which recruits EIF4A RNA helicase, as set forth in SEQ ID NO: 787, or an amino acid sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 97%, at least 98%, or at least 99% identical thereto.

[0224] In certain embodiments, the RNA helicase recruitment motif polypeptide is a canine hepatitis virus core protein motif polypeptide. In certain embodiments, the canine hepatitis virus core protein motif polypeptide (which recruits cellular DDX3) comprises an amino acid set forth in SEQ ID NO: 788, or an amino acid sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 97%, at least 98%, or at least 99% identical thereto. In certain embodiments, the canine hepatitis virus core protein motif polypeptide comprises an amino acid set forth in SEQ ID NO: 789, or an amino acid sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 97%, at least 98%, or at least 99% identical thereto. In certain embodiments, the canine hepatitis virus core protein motif polypeptide comprises an amino acid set forth in SEQ ID NO: 1853, or an amino acid sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 97%, at least 98%, or at least 99% identical thereto.

[0225] Without wishing to be bound by theory, fusing known RNA helicase domains to a RTE polypeptide, or fusing peptides that can recruit cellular RNA helicases to a RTE polypeptide, may improve reverse transcription in genome editing systems by removing secondary structure in the RNA template that may block reverse transcription thus reducing activity and processivity.

[0226] RAD51 is a protein involved in the homology directed repair (HDR) pathway, and a RAD51 domain directly binds to single-stranded DNA and searches for matching DNA in the genome as an early step in homologous recombination. Without being bound by theory, in site-specific transgene integration, e.g., transgene integration by R2 RTE proteins or certain embodiments of drivers comprising R2 RTE polypeptides, RAD51 fusion to a RTE polypeptide may improve integration at genomic sites that are homologous to sequences that are included on a transgene construct (for example sequences that flank the transgene and are homologous to one or more genomic sequences in the vicinity of the target gene).

[0227] In certain embodiments, the RAD51 polypeptide comprises an amino acid sequence set forth in SEQ ID NO: 21, or an amino acid sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 97%, at least 98%, or at least 99% identical thereto. In certain embodiments, the RAD17 polypeptide comprises an amino acid sequence set forth in SEQ ID NO: 34, or an amino acid sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 97%, at least 98%, or at least 99% identical thereto. In certain embodiments, the RAD6 polypeptide comprises an amino acid sequence set forth in SEQ ID NO: 27, or an amino acid sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 97%, at least 98%, or at least 99% identical thereto.

[0228] It will be appreciated that an engineered driver of the disclosure may comprise one or more, for example 1, 2, 3, 4, or 5, heterologous RNA/DNA processing polypeptides, optionally in combination with one or more other heterologous polypeptides as described herein.

[0229] In certain embodiments, the engineered driver protein comprises a R2 RTE polypeptide fused to one or more heterologous RNA/DNA processing polypeptides. In certain embodiments, the engineered driver protein comprises a R2 RTE polypeptide fused to a RNA helicase polypeptide (e.g., a EIF4A polypeptide or a Dengue virus RNA helicase polypeptide) or a RNA helicase recruitment motif polypeptide (e.g., a TUDOR domain polypeptide). The fused polypeptide (RNA helicase polypeptide or RNA helicase recruitment motif polypeptide) may be situated internally in the R2 RTE protein, on the N-terminal side of the R2 RTE polypeptide or the C-terminal side of the R2 RTE polypeptide. SEQ ID NOS:585 and 1875 set forth amino acid sequences of exemplary driver proteins comprising a RNA helicase polypeptide. SEQ ID

NOS:584 and 1853 set forth amino acid sequences of exemplary driver proteins comprising a RNA helicase recruitment motif polypeptide.

[0230] In certain embodiments, the engineered driver protein comprises a non-R2 RTE polypeptide fused to a RNA helicase polypeptide. In certain embodiments, the non-R2 RTE may be, e.g., Vingi RTE, R2NS RTE, L4_ACar RTE, a CR1_10_Ami RTE, a ZFL2-2 RTE, or any other non-R2 RTE disclosed elsewhere herein.

Heterologous localization signal

[0231] The engineered drivers of the disclosure may comprise heterologous localization sequences. For example, a nuclear localization sequence (NLS) is a polypeptide that promotes import of a protein into a nucleus of a cell. A nucleolar localization sequence (NoLS) is a polypeptide that promotes import of a protein into a nucleolus of a cell. NLS's and NoLS's, and methods for assessing their ability to direct a polypeptide to the nucleus or the nucleolus, respectively, are known to those with skill in the art.

[0232] In certain embodiments, the at least one heterologous polypeptide fused to a RTE polypeptide comprises a localization polypeptide. In certain embodiments, the localization polypeptide may be a nuclear localization signal (NLS). In certain embodiments, the localization polypeptide may be a nucleolar localization signal (NoLS).

[0233] In certain embodiments, the NLS comprises an amino acid sequence set forth in any one of SEQ ID NOs: 110-112, 116, or 1854. In certain embodiments, an NLS comprises a SV40 sequence (e.g., SEQ ID NO: 110), a nucleoplasmin sequence (e.g., SEQ ID NO: 111), a TOPBN1 sequence (e.g., SEQ ID NO: 112), or a bipartite SV40 sequence (e.g., SEQ ID NO: 116 or SEQ ID NO: 1854),

[0234] In non-limiting embodiments, the NoLS comprises an amino acid sequence set forth in any one of SEQ ID NOs: 113-115. In certain embodiments, a NoLS comprises a PNRC sequence (e.g., SEQ ID NO: 113), a poly R sequence (e.g., SEQ ID NO: 114), or a H2B sequence (e.g., SEQ ID NO: 115).

[0235] It will be appreciated that an engineered driver of the disclosure may comprise one or more, for example 1, 2, 3, 4, or 5, heterologous localization sequences, optionally in combination with one or more other heterologous polypeptides as described herein.

[0236] In certain embodiments, the engineered driver protein comprises a R2 RTE polypeptide fused to one or more heterologous localization sequences. In certain embodiments, the engineered driver protein comprises a R2 RTE polypeptide fused to a NLS or a NoLS. The fused polypeptide (NLS or NoLS) may be situated on the N-terminal side of the R2 RTE polypeptide or the C-terminal side of the R2 RTE polypeptide. SEQ ID NOS:636 and 716-718 set forth amino acid sequences of exemplary driver proteins comprising a NLS or NoLS.

Heterologous intrinsically disordered polypeptide

[0237] Intrinsically disordered regions (IDRs) are regions of proteins that lack a fixed three dimensional (3D) structure. These IDRs, which lack an amino acid sequence that form, for example, alpha helices, beta sheets, and other canonical 3D protein structures, are thought to provide flexibility, allowing the protein to have ordered interactions with multiple molecular partners within the intracellular environment.

[0238] IDRs also demonstrate a function of forming and maintaining condensates. Condensates can form and maintain organization through a range of different processes, the most well-known of which is phase separation of proteins and RNA. There are many different examples of intrinsically disordered proteins forming condensates. Condensation is also called phase-separation, and may provide beneficial properties. Without wishing to be bound by theory, condensed proteins might require or benefit from phase-separated environment with different and optimized chemical-physiological conditions or help with energetic limitations by reducing the entropy and increasing the enthalpy of the system. Without wishing to be bound by theory, phase separation can also prevent degradation, and help with nuclear import or to improve protein-RNA intermolecular interactions. For example, it has been shown that for LINE-1 elements, ORF1 tends to condensate. R2-TG has an intrinsically disordered region at its N-terminal (Residues 1-161, 1-184 or 1-218). Without wishing to be bound by theory, if a N-terminal intrinsically disordered region is responsible for/promotes condensation of R2 proteins, then other condensation peptides may be substituted for the disordered domains of RTE proteins and improve function.

[0239] In certain embodiments, the at least one heterologous polypeptide fused to a RTE polypeptide comprises a heterologous IDR polypeptide. In certain embodiments, the heterologous IDR polypeptide comprises an amino acid sequence set forth in any one of SEQ ID NOS: 1171-1216, or an amino acid sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 97%, at least 98%, or at least 99% identical thereto.

[0240] In some embodiments, the heterologous IDR polypeptide may be a condensation polypeptide comprising any one (or a combination of two or more) of the following: the N-terminal 161 residues of R2-TG protein (SEQ ID NO:1171); the N-terminal 184 residues of R2-TG RTE protein (SEQ ID NO:1172); the N-terminal 218 residues of R2-TG RTE protein (SEQ ID NO:1173), a condensation peptide from the first IDR of UBQ2 polypeptide (SEQ ID NO:1174), a condensation peptide from the second IDR of UBQ2 polypeptide (SEQ ID NO:1175), a condensation peptide from the third IDR of UBQ2 polypeptide (SEQ ID NO:1176), a condensation peptide from the second IDR of RBM14 polypeptide (SEQ ID NO:1177), a condensation peptide from the fourth IDR of RBM14 polypeptide (SEQ ID NO:1178), a condensation peptide from the first IDR of HSPB8 polypeptide (SEQ ID NO:1179), a condensation peptide from the first IDR of hnRNPA1 polypeptide (SEQ ID NO:1180), a condensation peptide from the second IDR of hnRNPA1 polypeptide (SEQ ID NO:1181), a condensation peptide from the third IDR of TAF15 polypeptide (SEQ ID NO:1182), a condensation peptide from the fourth IDR of TAF15 polypeptide (SEQ ID NO:1183), a condensation peptide from the first IDR of TDP43 polypeptide (SEQ ID NO:1184), a condensation peptide from the second IDR of TDP43 polypeptide (SEQ ID NO:1185), a condensation peptide from the first IDR of DDX4 polypeptide (SEQ ID NO:1186), a condensation peptide from the first IDR of FUS polypeptide (SEQ ID NO:1187), a condensation peptide from the second IDR of FUS polypeptide (SEQ ID NO:1188), a condensation peptide from the third IDR of FUS polypeptide (SEQ ID NO:1189), a condensation peptide from the first IDR of DYRK3 polypeptide (SEQ ID NO:1190), a condensation peptide from the first IDR of CWC27 polypeptide (SEQ ID NO:1191), a condensation peptide from the second IDR of CWC27 polypeptide (SEQ ID NO:1192), a condensation peptide from the first IDR of SPAG7 polypeptide (SEQ ID NO:1193), a condensation peptide from the second IDR of SPAG7 polypeptide (SEQ ID NO:1194), a condensation peptide from the first IDR of RS10L polypeptide (SEQ ID NO:1195), a condensation peptide from the first IDR of RS4Y2

polypeptide (SEQ ID NO:1196), a condensation peptide from the first IDR of RBMY1D polypeptide (SEQ ID NO:1197), a condensation peptide from the second IDR of RBMY1D polypeptide (SEQ ID NO:1198), a condensation peptide from the first IDR of KHDC4 polypeptide (SEQ ID NO:1199), a condensation peptide from the second IDR of KHDC4 polypeptide (SEQ ID NO:1200), a condensation peptide from the third IDR of KHDC4 polypeptide (SEQ ID NO:1201), a condensation peptide from the first IDR of RP9 polypeptide (SEQ ID NO:1202), a condensation peptide from the second IDR of RP9 polypeptide (SEQ ID NO:1203), a condensation peptide from the first IDR of PHP14 polypeptide (SEQ ID NO:1204), a condensation peptide from the first IDR of LMOD1a polypeptide (SEQ ID NO:1205), a condensation peptide from the second IDR of LMOD1a polypeptide (SEQ ID NO:1206), a condensation peptide from the first IDR of H2A1H polypeptide (SEQ ID NO:1207), a condensation peptide from the first IDR of CWC25 polypeptide (SEQ ID NO:1208), a condensation peptide from the first IDR of RAD51AP1 polypeptide (SEQ ID NO:1209), a condensation peptide from the first IDR of SPA24 polypeptide (SEQ ID NO:1210), a condensation peptide from the first IDR of HSB1L polypeptide (SEQ ID NO:1211), a condensation peptide from the first IDR of HIT polypeptide (SEQ ID NO:1212), a condensation peptide from the first IDR of RAMAC polypeptide (SEQ ID NO:1213), a condensation peptide from the first IDR of AIMP1 polypeptide (SEQ ID NO:1214), a condensation peptide from the first IDR of MRPL1 polypeptide (SEQ ID NO:1215), a condensation peptide from the second IDR of RAD51AP1 polypeptide (SEQ ID NO:1216). For clarity, the sequence identifiers provided above are the predicted IDRs of the polypeptides.

[0241] In some embodiments, a condensation peptide included in an engineered driver is capable of directly binding to the template RNA or to other condensation fused peptides. In some embodiments, in site-specific transgene integration, condensation peptide fusion to a RTE polypeptide may improve integration at genomic sites that are homologous to sequences that are included on a transgene construct (for example sequences that flank the transgene and are homologous to one or more genomic sequences in the vicinity of the target gene, e.g., homology arms). In general, the transgene may be in sense or anti-sense orientation.

[0242] Many RTE proteins have native IDRs. For example, as noted above, R2 RTE proteins typically have a native IDR at its N-terminal region. In certain embodiments, the RTE

polypeptide to which the heterologous IDR polypeptide is fused is a RTE polypeptide having a deletion of its native IDR or portion thereof.

[0243] It will be appreciated that an engineered driver of the disclosure may comprise one or more, for example 1, 2, 3, 4, or 5, heterologous IDR polypeptides, optionally in combination with one or more other heterologous polypeptides as described herein.

[0244] In certain embodiments, the engineered driver protein comprises a R2 RTE polypeptide fused to one or more heterologous IDR polypeptides. In certain embodiments, the engineered driver protein comprises a R2 RTE polypeptide fused to a heterologous IDR polypeptide. The fused heterologous IDR polypeptide be situated internally in the R2 RTE protein, on the N-terminal side of the R2 RTE polypeptide or the C-terminal side of the R2 RTE polypeptide. SEQ ID NOS: 679-714, 775-781 set forth amino acid sequences of exemplary driver proteins comprising a heterologous IDR polypeptide.

[0245] In certain embodiments, the engineered driver protein comprises a non-R2 RTE polypeptide fused to one or more heterologous IDR polypeptides. In certain embodiments, the non-R2 RTE may be, e.g., Vingi RTE, R2NS RTE, L4_ACar RTE, a CR1_10_Ami RTE, a ZFL2-2 RTE, or any other non-R2 RTE disclosed elsewhere herein.

Heterologous endonuclease – Chimeric drivers

[0246] In certain embodiments, the at least one heterologous polypeptide fused to a RTE polypeptide comprises a heterologous endonuclease (EN) domain polypeptide. As described herein, a RTE protein typically comprises an EN domain and an RT domain. As such, a heterologous EN domain polypeptide in the context of engineered drivers as described herein refers to an EN domain polypeptide of a protein that is not the RTE protein from which the RTE polypeptide (which may comprise the RT domain) of the engineered driver is derived. As such, the heterologous EN domain may be a EN domain polypeptide of a different RTE protein, or a EN domain polypeptide of a non-RTE protein.

[0247] For example, if a given engineered driver comprises a RTE polypeptide of a R2-1-TG RTE protein, then the heterologous EN domain may be a EN domain polypeptide of a different R2 protein (e.g., a R2-2-AcOr RTE protein or a R4-AC RTE protein). In such a case, the

engineered driver would be a chimeric driver protein comprising EN and RT domains derived from different R2 proteins.

[0248] In certain embodiments, if a given engineered driver comprises a RTE polypeptide of a R2 protein, then the heterologous EN domain polypeptide may be a EN domain polypeptide of a non-R2 RTE protein. In such a case, the engineered driver would be a chimeric driver protein comprising EN and RT domains derived from different RTE proteins.

[0249] In certain embodiments, if a given engineered driver comprises a RTE polypeptide of a RTE (e.g. R2 RTE) protein, then the heterologous EN domain may be the EN domain of a non-RTE protein. In such a case, the engineered driver would be a chimeric driver protein comprising an EN domain from a non-RTE source. Examples of non-RTE proteins comprising a EN domain include but are not limited to a Cas nuclease (e.g., SpCas9 nuclease), a Cas9 nickase (e.g., SpCas9 with a H840 mutation), a homing endonuclease, or a FokI nuclease. Fanzor nuclease (e.g., SpuFz1) is a RNA-guided system (aka OMEGA). Like the Cas systems, it uses a guide RNA molecule to guide the binding and cleavage of specific and non-specific dsDNA sequences. In certain embodiments, the heterologous EN domain includes but is not limited to a Fanzor1 or Fanzor2 nuclease (e.g., a Fanzor1 nuclease or Fanzor2 dead nuclease). Non-limiting examples of drivers utilizing a Fanzor1 or Fanzor2 nuclease have amino acid sequence set forth in any one of SEQ ID NOS: 160-172.

[0250] In certain embodiments, the heterologous EN domain polypeptide may functionally replace the native EN activity of the RTE polypeptide of the engineered driver. In certain embodiments, the RTE polypeptide of the engineered driver may be lacking in a functional EN domain, for example through deletion of the EN domain or portion thereof, and/or through one or more amino acid substitutions that reduce or eliminate enzymatic activity of the native EN domain of the RTE polypeptide.

[0251] Examples of chimeric drivers with a heterologous EN domain polypeptide provided herein include at least the following: SEQ ID NO: 133 sets forth the amino acid sequence of a chimeric driver of different R2 proteins (R2-ZA and R2-TG), comprising a R2-TG RTE polypeptide and a R2-ZA EN domain (which replaced the R2-TG EN domain). SEQ ID NO: 160 sets forth the amino acid sequence of a SpuFz1 nuclease R2-TG fusion. The SpuFz1 portion is fused at the N-terminus via a linker to a C-terminal R2-TG protein. SEQ ID NO: 161 sets forth

the amino acid sequence of a SpuFz1 nuclease R2-TG fusion. The SpuFz1 portion is fused at the N-terminus via a linker to a truncated C-terminal R2-TG protein, lacking first 209 amino acid residues. SEQ ID NO: 162 sets forth the amino acid sequence of a SpuFz1 nuclease R2-TG fusion. The SpuFz1 portion is fused at the N-terminus via a linker to a truncated C-terminal R2-TG protein, with native endonuclease domain deleted. SEQ ID NO: 163 sets forth the amino acid sequence of a SpuFz1 nuclease R2-TG fusion. The SpuFz1 portion is fused at the C-terminus via a linker to a truncated N-terminal R2-TG protein, with native endonuclease domain deleted. SEQ ID NO: 164 sets forth the amino acid sequence of a SpuFz1 nuclease R2-TG fusion. The SpuFz1 portion is fused at the C-terminus via a linker to an N-terminal R2-TG protein, carrying a K1307A mutation. SEQ ID NO: 165 sets forth the amino acid sequence of a dead SpuFz1 nuclease R2-TG fusion. The SpuFz1 carries a D606A mutation and is fused at the N-terminus fused via a linker to a C-terminal R2-TG protein. SEQ ID NO: 166 sets forth the amino acid sequence of a dead SpuFz1 nuclease R2-TG fusion. The SpuFz1 carries a D606A mutation and is fused at the N-terminus via a linker to a C-terminal R2-TG protein, lacking first 209 amino acid residues. SEQ ID NO: 167 sets forth the amino acid sequence of a SpuFz1 nuclease R2-TG fusion. The SpuFz1 is fused at the N-terminus via a linker to a C-terminal R2-TG protein (lacking 209 N-terminal amino acids. SEQ ID NO: 168 sets forth the amino acid sequence of a SpuFz1 nuclease R2-TG fusion. The SpuFz1 is fused at the N-terminus via a linker to a C-terminal R2-TG protein without the native DNA binding domain, and carrying K1307A mutation. SEQ ID NO: 169 sets forth the amino acid sequence of a SpuFz1 nuclease R2-TG fusion. The SpuFz1 is fused at the N-terminus via a linker to a C-terminal R2-TG protein carrying a K1307A mutation. SEQ ID NO: 170 sets forth the amino acid sequence of a SpuFz1 nuclease R2-TG fusion. The SpuFz1 is fused at the C-terminus via a linker to an N-terminal R2-TG protein (lacking 209 N-terminal amino acid residues). SEQ ID NO: 171 sets forth the amino acid sequence of a SpuFz1 nuclease R2-TG fusion. In each of the above, the SpuFz1 is fused at the C-terminus via a linker to a truncated N-terminal R2-TG protein.

Heterologous reverse transcriptase domain

[0252] In certain embodiments, the at least one heterologous polypeptide fused to a RTE polypeptide comprises a heterologous reverse transcriptase (RT) domain polypeptide.

[0253] As described herein, a RTE protein typically comprises an EN domain and an RT domain. As such, a heterologous RT domain polypeptide in the context of engineered drivers as

described herein refers to an RT domain polypeptide of a protein that is not the RTE protein from which the RTE polypeptide (which may comprise the EN domain) of the engineered driver is derived. As such, the heterologous RT domain polypeptide may be a RT domain polypeptide of a different RTE protein, or a RT domain polypeptide of a non-RTE protein.

[0254] For example, if a given engineered driver comprises a RTE polypeptide of a R2-1-TG RTE protein, then the heterologous RT domain polypeptide may be a RT domain polypeptide of a different R2 protein (e.g., a R2-2-AcOr RTE protein or a R4-AC RTE protein). In such a case, the engineered driver would be a chimeric driver protein comprising EN and RT domains derived from different R2 proteins. In certain embodiments, the RT domain polypeptide of a different R2 RTE protein may be derived from a R2-Toc RT domain, optionally comprising an amino acid sequence set forth in SEQ ID NO: 295, or an amino acid sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 97%, at least 98%, or at least 99% identical thereto.

[0255] In certain embodiments, if a given engineered driver comprises a RTE polypeptide of a R2 protein, then the heterologous RT domain polypeptide may be a RT domain polypeptide of a non-R2 RTE protein. In such a case, the engineered driver would be a chimeric driver protein comprising EN and RT domains derived from different RTE proteins.

[0256] In certain embodiments, the RT domain polypeptide of a non-R2 RTE protein may be from an LTR retroelement or an ERV. In certain embodiments, the RT domain polypeptide of a non-R2 RTE protein may be from a non-LTR retroelement. In certain embodiments, the RT domain polypeptide of a non-R2 RTE protein may be from a Vingi RTE. In certain embodiments, the RT domain polypeptide of a non-R2 RTE protein may be from a R2NS RTE. In certain embodiments, the RT domain polypeptide of a non-R2 RTE protein may be from a L2_L4_ACar RTE. In certain embodiments, the RT domain polypeptide of a non-R2 RTE protein may be from a CR1_10_Ami RTE. In certain embodiments, the RT domain polypeptide of a non-R2 RTE protein may be from a ZFL2-2 RTE.

[0257] In certain embodiments, if a given engineered driver comprises a RTE polypeptide of a RTE (e.g. R2 RTE) protein, then the heterologous RT domain polypeptide may be the RT domain polypeptide of a non-RTE protein. In such a case, the engineered driver would be a chimeric driver protein comprising an RT domain from a non-RTE source. In certain

embodiments, the RT domain comprises a sequence may be from a MMLV RT, Marathon RT, a TGIRT_III RT, or a SSIV RT.

[0258] In certain embodiments, the heterologous RT domain polypeptide may functionally replace the native RT activity of the RTE polypeptide of the engineered driver. In certain embodiments, the RTE polypeptide of the engineered driver may be lacking in a functional RT domain, for example through deletion of the RT domain or portion thereof, and/or through one or more amino acid substitutions that damage or eliminate enzymatic activity of the native RT domain of the RTE polypeptide.

[0259] SEQ ID NO: 290 sets forth an amino acid sequence of a Vingi RT. SEQ ID NO: 291 sets forth amino acid sequence of an MMLV RT. SEQ ID NO: 292 sets forth amino acid sequence of a R2NS RT. SEQ ID NO: 293 sets forth amino acid sequence of a L2_L4_ACar RT. SEQ ID NO: 294 sets forth amino acid sequence of a CR1-10_Ami RT. SEQ ID NO: 295 sets forth amino acid sequence of a R2 Toc RT. SEQ ID NO: 296 sets forth amino acid sequence of a ZFL2-2 RT. SEQ ID NO: 297 sets forth amino acid sequence of a Marathon RT. SEQ ID NO: 298 sets forth amino acid sequence of a TGIRT_III RT. SEQ ID NO: 299 sets forth amino acid sequence of an MMLV derived RT peptide. SEQ ID NO: 300 sets forth amino acid sequence of a SSIV RT.

[0260] Examples of chimeric drivers with a heterologous RT domain polypeptide provided herein include at least the following: R2-TG chimera with Vingi RT (SEQ ID NO: 270), R2-TG chimera with MMLV RT (SEQ ID NO: 271), R2-TG chimera with R2NS RT (SEQ ID NO: 272), R2-TG chimera with L2_24_ACar-RT (SEQ ID NO: 273), R2-TG chimera with CR1-10_Ami-RT (SEQ ID NO: 274), R2-TG chimera with R2Toc-RT (SEQ ID NO: 275), R2-TG chimera with ZFL2-2 RT (SEQ ID NO: 276), R2-TG chimera with Marathon RT (SEQ ID NO: 277), R2-TG chimera with TGIRT_III RT (SEQ ID NO: 278), R2-TG mutated RT with YADD RT motif (SEQ ID NO: 279), R2-TG with RT deletion (SEQ ID NO: 280), R2-TG internal fusion with MMLV RT derived peptide in the RT domain (SEQ ID NO: 281), R2-TG chimera with mutated R2-TG and MMLV RT sequence from SEQ ID NO: 281 (SEQ ID NO: 282), R2-TG chimera with SSIV RT (SEQ ID NO: 283), mutated R2-TG fusion with SSIV RT (SEQ ID NO: 284), R2-TG internal fusion with N-terminal fused RNA binding domain (RBD) motif (SEQ ID NO: 285), R2-TG internal fusion of RBD motif (SEQ ID NO: 286). SEQ ID NO: 134 sets forth the amino acid sequence of a chimeric driver between different R2 proteins (R2-ZA

and R2-TG), comprising a R2-TG RTE polypeptide and a R2-ZA RT domain (which replaced the R2-TG RT domain).

Specific combinations of heterologous polypeptides

[0261] Certain combinations of heterologous polypeptides were found to provide unexpectedly advantageous properties in the resulting engineered drivers, for example as demonstrated in the Examples below. These advantageous combination include, but are not limited to, an engineered driver, e.g., a driver comprising a R2 polypeptide, fused to a CTIP polypeptide and a HMGN1 polypeptide, as well as an engineered driver, e.g., a driver comprising a R2 polypeptide, fused to a Sto7d polypeptide and a HMGN1 polypeptide.

[0262] In certain embodiments, the engineered driver comprises a R2 polypeptide fused to a CTIP polypeptide and a HMGN1 polypeptide. In certain embodiments, the engineered driver comprises, in an N-terminal to C-terminal order, the CTIP polypeptide, the HMGN1 polypeptide, and the R2 polypeptide. In certain embodiments, the R2 polypeptide may be lacking in a native N-terminal intrinsically disordered region.

[0263] In certain embodiments, the engineered driver comprises a R2 polypeptide fused to a CTIP polypeptide, a HMGN1 polypeptide, and a heterologous intrinsic disordered region. In certain embodiments, the engineered driver comprises, in an N-terminal to C-terminal order, a heterologous intrinsic disordered region, a CTIP polypeptide, a HMGN1 polypeptide, and a R2 polypeptide. In certain embodiments, the R2 polypeptide may be lacking in a native N-terminal intrinsically disordered region. By way of example, a driver polypeptide having the above-noted arrangement of components may comprise an amino acid sequence set forth in any one of SEQ ID NOS: 679-714

[0264] In certain embodiments, the engineered driver comprises a R2 polypeptide fused to a CTIP polypeptide, a HMGN1 polypeptide, and a RNA helicase polypeptide. In certain embodiments, the engineered driver comprises, in an N-terminal to C-terminal order, a RNA helicase polypeptide, a CTIP polypeptide, a HMGN1 polypeptide, and a R2 polypeptide. The RNA helicase polypeptide may be, e.g., a EI4A polypeptide or a Dengue virus RNA helicase polypeptide. In certain embodiments, the R2 polypeptide may comprises a deletion within its native N-terminal intrinsically disordered region. By way of example, a driver polypeptide

having the above-noted arrangement of components may comprise an amino acid sequence set forth in SEQ ID NO: 783.

[0265] In certain embodiments, the engineered driver comprises a R2 polypeptide fused to a Sto7d polypeptide and a HMGN1 polypeptide. In certain embodiments, the engineered driver comprises, in an N-terminal to C-terminal order, the Sto7d polypeptide, the HMGN1 polypeptide, and the R2 polypeptide. In certain embodiments, the R2 polypeptide may be lacking in a native N-terminal intrinsically disordered region.

[0266] In certain embodiments, the engineered driver comprises a R2 polypeptide fused to a Sto7d polypeptide, a HMGN1 polypeptide, and a RNA helicase recruitment motif polypeptide. In certain embodiments, the engineered driver comprises, in an N-terminal to C-terminal order, the RNA helicase recruitment motif polypeptide, the Sto7d polypeptide, the HMGN1 polypeptide, and the R2 polypeptide. The RNA helicase recruitment motif polypeptide may be, e.g., a TUDOR domain polypeptide. In certain embodiments, the R2 polypeptide may comprises a deletion within its native N-terminal intrinsically disordered region. By way of example, a driver polypeptide having the above-noted arrangement of components may comprise an amino acid sequence set forth in SEQ ID NO: 1234.

Modifications made to RTE proteins or portions thereof

[0267] In certain embodiments, a RTE polypeptide comprises at least one amino acid substitution, at least one amino acid deletion, at least one amino acid addition, or a combination thereof, when compared to a wildtype (WT) RTE protein.

[0268] In certain embodiments, the RTE polypeptide may have least 70%, at least 71%, at least 72%, at least 73%, at least 74%, at least 75%, at least 76%, at least 77%, at least 78%, at least 79%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% sequence identity with the WT RTE protein.

Amino acid substitution

[0269] In certain embodiments, a RTE polypeptide comprises at least one amino acid substitution, relative to the corresponding WT RTE protein.

[0270] In certain embodiments, an amino acid substitution (e.g., in a reverse transcriptase domain or an endonuclease domain) is a stabilizing amino acid substitution, for example a substitution that improves packing of hydrophobic residues in the core of the domain, stabilizes a loop region, and/or alters electrostatic, H-bond stability, or S-bond stability. In certain embodiments, the substitution and/or addition that stabilizes a loop region is a proline substitution. In certain embodiments, the substitution that alters electrostatic, H-bond stability, or S-bond stability adds a positive charge, for example by substitution to lysine or arginine, or substitution from aspartate or glutamate to a non-charged residue such as alanine.

[0271] In certain embodiments, the at least one amino acid substitution stabilizes the reverse transcriptase domain of the R2 protein and/or the association of the reverse transcriptase domain with RNA and/or DNA, relative to the reverse transcriptase domain of the WT of the R2 protein.

[0272] In certain embodiments, the at least one amino acid substitution is relative to SEQ ID NO: 1. In certain embodiments, the at least one amino acid substitution corresponds to at least one substitution (e.g., 1, 2, 3, 4, 5, etc. substitutions) selected from E399M, Y477H, D555K, W762R, R822K, H1090Y, F956I, P78V, W677C, K314R, K385M, P437M, T448Y, C518Y, T305R, P280K, N269R, L213K, E214K, Q215K, E214R, T1011R, K547R, C145G, I1219N, G257M, K700D, F148K, F667G, and F667Y relative to SEQ ID NO: 1. In various embodiments, the at least one amino acid substitution corresponds to at least one substitution (e.g., 1, 2, 3, 4, 5, etc. substitutions) selected from E399M, Y477H, D555K, W762R, R822K, H1090Y, F956I, P78V, W677C, K314R, K385M, P437M, T448Y, C518Y, T305R, P280K, N269R, L213K, E214K, Q215K, E214R, and T1011R relative to SEQ ID NO: 1. In certain embodiments, the at least one amino acid substitution is a substitution corresponds to at least one of positions F148, K700, F667, C145, P280, D555, T1011 relative to SEQ ID NO:1. In certain embodiments, the at least one amino acid substitution may be a substitution corresponding to position F148 relative to SEQ ID NO:1. In certain embodiments, the at least one amino acid substitution may be a substitution corresponding to position K700 relative to SEQ ID NO:1. In certain embodiments, the at least one amino acid substitution may be a substitution

corresponding to position F667 relative to SEQ ID NO:1. In certain embodiments, the at least one amino acid substitution may be a substitution corresponding to position P280 relative to SEQ ID NO:1. In certain embodiments, the at least one amino acid substitution may be a substitution corresponding to position D555 relative to SEQ ID NO:1. In certain embodiments, the at least one amino acid substitution may be a substitution corresponding to position T1011 relative to SEQ ID NO:1. In certain embodiments, the at least one amino acid substitution may be a substitution corresponding to position K700 relative to SEQ ID NO:1. In certain embodiments, the retroelement polypeptide is a R2-TG polypeptide, which may, e.g., have an amino acid sequence set forth in any one of SEQ ID NOs: 230-238.

[0273] In certain embodiments, a reverse transcriptase domain of the RTE polypeptide is a variant reverse transcriptase domain, relative to a natively occurring unsubstituted RT domain. In certain embodiments, a variant reverse transcriptase domain comprises at least one amino acid substitution that improves at least one of stability, interaction with RNA, or interaction with DNA relative to an unsubstituted reverse transcriptase domain. In certain embodiments, the variant reverse transcriptase domain comprises at least one amino acid substitution that stabilizes its association with RNA or DNA relative to an unsubstituted reverse transcriptase domain. In certain embodiments, the at least one amino acid substitution adds a positive charge (e.g., via the addition of lysine or arginine), removes a negative charge (e.g., via the removal of an aspartate or glutamate), alters at least one H-bond forming residue, or alters at least one S-bond forming residue. In certain embodiments, the at least one amino acid substitution corresponds to a substitution in the RTE reverse transcriptase domain. In certain embodiments, the at least one amino acid substitution corresponds to a substitution in the RTE endonuclease domain. In certain embodiments, the at least one amino acid substitution corresponds to a substitution in the DNA or RNA binding domains. In certain embodiments, the at least one (in some embodiments 1, 2, 3, etc.) amino acid substitution may be selected from: L18P, E19L, S26N, V27R, A29T, I30A, R32G, N35I, S36R, L37K, A38R, L49F, G209R, K289R, L316H, E399M, E472G, Y477H, T514R, D537Y, T538K, H544Q, D555K, N562D, W762R, R822K, S895K, Q909N, R964K, G966N, T989Y, I1003M, T1011S, A1012G, L1013R, T1016L, Q1019R, E1027D, P1033R, C1079K, K1085R, L1089R, H1090Y, I1219N, A1282G, T989I, K579R, K547R, V660I, T608P, F956I, I1003L, T1016S, T680S, L594P, P52R, P418S, P78G, L51G, P78V, P180S, V539G, P180K, P437R, V539R, N640A, W677C, V687W, R866I, A747G, W762E, T908I, W959K, D1037N, Y1194L, R1292P, K1210R, K993R, K755R, C518R, K314R,

K1182R, K385M, P437M, T448Y, C518Y, E1008R, T305R, P280K, A1012R, N269R, L213K, E214K, Q215K, E214R, T1011R, E1015R, L213R, G257M, G209F, W762D, W959K, T908I, Y579R, T1011A, Y571H, E630D, D1075E, L1089M, A442P, H423Q, E173D, P188L, C145G, N301G, R355K, G257R, V251G, T603A, N363H, P1033S, F667Y, F1365L, V501L, G257R+C145G, W762R+C145G, G257R+E173D, N363H+E173D, A442P+E173D, W762R+E173D, G257R+P188L, W762R+P188L, G257R+V251G, S895K+V251G, N363H+G257R, R568S, E542K, P985R, G561R, Y563R, L1184K, H1030R, A557R, I558R, Q567R, N271W, E350R, L316R, F304K, D923K, A317W, E1010R, Q1021R, E419R, T1011F, E769R, W345R, V110R, R568G, F135G, N104Y, V142Y, P115K, A111Y, L129R, V98R, D126R, E144R, Q112R, Y136R, F148K, T1016K, K700D, E188D, E188A, R435K, L438A, A1191L, P1195G, P1195G, R1197K, L1200I, G1331A, G1335A, P593G, S595V, E596S, S599V, P593A, S595A, E596A, S599A, K612D, D616K, K612A, D616A, K856D, D6860K, K856A, D6860A, N891S, T892V, S895V, N891A, T892A, S895A, E888S, E898S, T899V, N902S, E888A, E898A, T899A, N902A, T934V, G937P, P939G, T934A, G937A, K1252D, D1255K, W1256G, K1252A, D1255A, W1256A, A877V, A877I, A877M, F876Y, A877H, Q955R, L952I, T989L, E947S, G658R, Q642R, F973R, G647T, D990V, D990R, N941K, V727L, M874L, D227W, D227F, C344V, R669V, V600T, R866L, T989V, M449E, T679L, Y478L, M671S, F956V, M686S, W1126P, T1367R, C404S, E1363K, F973Q, M420T, R669F, G689E, N1219L, K480R, K508Q, E812K, V1368R, R1157M, G217P, H1320E, N891K, K1269G, K1269G, D1255G, T670L, T670L, H1180Q, K907R, T996A, I778F, R574G, K1151H, E194G, R574S, D1288A, F186G, T969E, G432Q, D591Q, M553R, T174K, W1028K, D534Y, D591W, D972R, W1028Y, D972Q, P169R, L975R, C587K, I978K, D591R, D182R, D972M, A1024K, C170K, P985K, G172R, S178R, G172K, S177R, A166R, S177W, P548R, S204R, H544R, G575H, F1135T, N1123K, P1116D, P998H, S645Q, P998Q, D879A, D879M, K576M, D879W, E643R, T994W, N1140R, D1133R, D583R, M644R, G575R, A1128R, A649R, E1125R, F1135Y, C750K, P1129R, F1135R, L1057R, H1006K, L582K, D573R, T994R, D878R, A578K, D839R, G311S, T532K, T728K, P1116K, L243K, E247K, L249F, K255R, V258I, F260Y, Y272H, E286K, P291A, D303S, I308R, I327L, K333R, G339R, I370L, I374L, K377R, S383K, L474I, T481I, L482I, F498L, M502L, Q516R, F519I, G520K, C521L, S526K, D534K, L572K, S589R, F611Y, P545K, F549R, L564R, L569R, F570Y, D573N, K377R, L582I, I585K, I597V, F601Y, F628L, L631P, V641L, A649S, P650A, V661K, E666K, E672R, F674Y, W677I, V687L, C690S, N738H, Q741R, L754I, A764S, H768R, L771I, V776I,

F846Y, L848I, I896L, W933F, Y995F, A1038G, A1063L, I778F, V785I, L794M, E798G, K821G, K821R, V835I, L844I, L854I, C855R, H865K, C901L, D929N, A932P, M942L, K950R, Y951H, D957N, S965P, T969E, P998S, Y1002H, L1014I, T1023K, E1027K, Y1041H, S1043R, K1053R, V1061R, Q1062R, T1130K, Q1173R, Q1233K, F1278L, V1287I, V1291I, E1317Q, V1318I, M1361L, E1363K, R1064K, K1078R, F1080L, E1082K, K1092R, G1098R, E1127K, E1127Q, K1132R, Q1155H, F1162Y, W1171F, L1184I, Q1188K, C1215Y, N1228Q, E1249K, V1257T, H1313P, F1345N, K1355R, K1386R, V727M, T759G, Q795R, I814C, T819E, H825K, Q868G, Q868D, A931P, D944G, F973S, L175S, A1012S, I683L, S699G, I707P, W710Y, T714A, I718T, I726V, N813G, V881T, L906M, S648K, S699K, S717P, A748R, D800P, V239M, E334A, T532G, K622Q, T1367R, H1378R, F1380Y, Y817R, Q911K, Q911S, L657Y, L720Y, A779S, K780N, D199K, A779K, D783G, D851N, E949Y, E949Y, V299E, E949Y, G520A, D591G, S599D, T1122L, L1285I, N1103D, P739R, E638L, N889G, Y1242K, E400M, D1133P, S645K, V639T, I745R, G772A, A747S, A748P, I778F, M841L, M850L, E949L, F983E, R669L, L213K, L564Q, F973E, R669I, F667G, and F973E, or combinations thereof, relative to SEQ ID NO: 1.

[0274] In certain embodiments, the at least one (in some embodiments 1, 2, 3, etc.) amino acid substitution may be selected from E399M, Y477H, D555K, W762R, R822K, H1090Y, F956I, P78V, W677C, K314R, K385M, P437M, T448Y, C518Y, T305R, P280K, N269R, L213K, E214K, Q215K, T1011R, R435K, L438A, R1197K, G1331A, T989L, Q955R, L952I, E947S, Q642R, G647T, N941K, V727L, M874L, D227W, D227F, C344V, R669V, V600T, T989V, M449E, T679L, Y478L, M671S, F956V, M686S, W1126P, T1367R, C404S, E1363K, K508Q, E812K, V1368R, R1157M, N1123K, P1116D, P998H, S645Q, P998Q, N1123K, P1116D, P998H, S645Q, P998Q, D879A, D879M, K576M, G311S, T532K, L572K, S589R, F611Y, F846Y, L848I, I896L, W933F, Y995F, A1038G, A1063L, T1130K, Q1173R, Q1233K, F1278L, V1287I, V1291I, E1317Q, V1318I, M1361L, E1363K, K1386R, V727M, T759G, Q795R, I814C, T819E, H825K, Q868G, Q868D, A931P, D944G, F973S, L175S, A1012S, I683L, S699G, I707P, W710Y, T714A, I718T, I726V, N813G, V881T, L906M, S648K, S699K, S717P, A748R, D800P, V239M, E334A, T532G, K622Q, D923K, P739R, E638L, N889G, Y1242K, E400M, T1011S, T1011A, I1219N, A1282G, L213R, A1012G, E542K, G257R, C145G, L316R, D537Y, D1133P, and K1085R, or combinations thereof, relative to SEQ ID NO: 1.

[0275] In certain embodiments, the at least one (in some embodiments 1, 2, 3, etc.) amino acid substitution may be selected from E399M, Y477H, D555K, W762R, R822K, H1090Y, F956I, P78V, W677C, K314R, K385M, P437M, T448Y, C518Y, T305R, P280K, N269R, L213K, E214K, Q215K, E214R, T1011R, K547R, C145G, I1219N, G257M, K700D, F148K, F667G, and F667Y, or combinations thereof, relative to SEQ ID NO:1.

[0276] In certain embodiments, the at least one amino acid substitution comprises a substitution corresponding to at least one ((in some embodiments 1, 2, 3, etc.) of positions F148, K700, F667, C145, P280, D555, T1011, I1219, and A1282 relative to SEQ ID NO:1.

[0277] In certain embodiments, the at least one amino acid substitution comprises a substitution corresponding to a substitution in position P280 relative to SEQ ID NO:1. The substitution may be selected from, for example, P280K, P280R, P280Q, P280N, and P280H.

[0278] In certain embodiments, the at least one amino acid substitution comprises a substitution corresponding to a substitution in position D555 relative to SEQ ID NO:1. The substitution may be selected from, for example, D555K, D555R, D555Q, D555A, D555H, D555S, D555C, D555Y, D555W, D555F, D555L, D555V, D555M, and D555I relative to SEQ ID NO: 1. In certain embodiments, the at least one amino acid substitution comprises the substitution of D555K relative to SEQ ID NO: 1. SEQ ID NO: 230 sets forth the mutated amino acid sequence of an exemplary R2-TG driver with the D555K amino acid substitution.

[0279] In certain embodiments, the at least one amino acid substitution comprises a substitution corresponding to a substitution in position T1011. T The substitution may be selected from, for example, to T1011R, T1011A, T1011K, T1011S, T1011F, T1011Y, T1011W, and T1011H relative to SEQ ID NO: 1. In certain embodiments the amino acid substitution is in position F148. The substituted amino acid may be selected from, for example, F148R, F148A, F148K, F148S, F148F, F148Y, F148W, F148H. In certain embodiments the amino acid substitution is in position K700. The substituted amino acid may be selected from, for example, K700D, K700E, K700S, K700Q, K700T, K700R, K700H.

[0280] In certain embodiments, the at least one amino acid substitution comprises at least the following substitutions: P280K, D555K, T1011S, I1219N, and A1282G relative to SEQ ID NO:1. Examples of engineered drivers comprising the P280K, D555K, T1011S, I1219N, and

A1282G substitutions relative to SEQ ID NO:1 are provided in SEQ ID NOS: 636, 633, 1236, 1234, 783, and 3028.

[0281] In certain embodiments, the at least one (e.g. 1, 2, 3, 4, 5, etc.) amino acid substitution comprises one or more of the following substitutions: G257R, P280K, N363H, E399M, V539G, H423Q, A442P, D537Y, K547R, D555K, R669L, N889G, F973E, T1011A, Q1019R, D1133P, K1085R, R1177G, Y1242K, and A1282G, or combinations thereof, relative to SEQ ID NO:1. In certain embodiments, the at least one amino acid substitution comprises at least the following substitutions: G257R, P280K, N363H, E399M, V539G, H423Q, A442P, D537Y, K547R, D555K, R669L, N889G, F973E, T1011A, Q1019R, D1133P, K1085R, R1177G, and Y1242K, A1282G relative to SEQ ID NO:1. An example of an engineered driver comprising the substitutions of G257R, P280K, N363H, E399M, V539G, H423Q, A442P, D537Y, K547R, D555K, R669L, N889G, F973E, T1011A, Q1019R, D1133P, K1085R, R1177G, Y1242K, and A1282G, relative to SEQ ID NO:1 is provided in SEQ ID NO: 668.

[0282] Examples of engineered drivers comprising at least one amino acid substitution relative to SEQ ID NO: 1 are provided in SEQ ID NOS: 230-238, 668 and 1378.

[0283] In certain embodiments, the at least one amino acid substitution is relative to SEQ ID NO: 2.

[0284] In certain embodiments, the at least one (in some embodiments 1, 2, 3, etc.) amino acid substitution may be selected from G1057D, S931R, A289P, V996S, L1100F, M287A, C482S, M685L, S547P, V916I, A1256E, M1259A, N332G, M358L, Y919P, T950A, M376P, P151K, D584L, Y1162Q, M640S, C314L, R1131G, A965S, Y1044L, D1236G, A970L, Q1039R, C474Y, and T404, , or combinations thereof, relative to SEQ ID NO: 2.

[0285] In certain embodiments, the at least one amino acid substitution comprises a substitution corresponding to at least one of positions S931, Q1039 or M287 relative to SEQ ID NO:2.

Amino acid deletion in RTE protein

[0286] In certain embodiments, the RTE polypeptide comprises at least one amino acid deletion. The deletion can be a deleting of individual amino acids, or a longer segment, including deletion of functional domains, binding domains, motifs, protein regions, or portions thereof. Segments

of a RTE protein that may be deleted in a RTE polypeptide may be, a RT domain, EN domain, a DNA binding domain, a RNA binding domain, a PIP motif, an intrinsically disordered region (IDR), or combinations thereof.

Deletion of N-terminal intrinsically disordered region

[0287] The N-terminal region of some RTE proteins, e.g., R2 proteins encoded in Group 2 RTE, contain intrinsically disordered regions (IDRs), which lack a fixed three dimensional (3D) structure. These IDRs are thought to provide flexibility, allowing the protein to have ordered interactions with multiple molecular partners within the intracellular environment. In certain embodiments, a RTE polypeptide comprised in an engineered driver protein may comprise a deletion of a N-terminal intrinsically disordered region (IDR) or a portion thereof of the corresponding WT RTE protein (e.g., a WT R2 RTE protein).

[0288] In certain embodiments, the deletion is a deletion of between about 10 amino acids and about 300 amino acids, between about 20 amino acids and about 300 amino acids, between about 40 amino acids and about 300 amino acids, between about 50 amino acids and about 300 amino acids, between about 60 amino acids and about 300 amino acids, between about 80 amino acids and about 300 amino acids, between about 90 amino acids and about 300 amino acids, between about 100 amino acids and about 300 amino acids, between about 120 amino acids and about 300 amino acids, between about 140 amino acids and about 300 amino acids, between about 150 amino acids and about 300 amino acids, between about 160 amino acids and about 300 amino acids, between about 180 amino acids and about 300 amino acids, between about 200 amino acids and about 300 amino acids, between about 220 amino acids and about 300 amino acids, between about 250 amino acids and about 300 amino acids, between about 10 amino acids and about 250 amino acids, between about 20 amino acids and about 250 amino acids, between about 40 amino acids and about 250 amino acids, between about 50 amino acids and about 250 amino acids, between about 60 amino acids and about 250 amino acids, between about 80 amino acids and about 250 amino acids, between about 90 amino acids and about 250 amino acids, between about 100 amino acids and about 250 amino acids, between about 120 amino acids and about 250 amino acids, between about 140 amino acids and about 250 amino acids, between about 150 amino acids and about 250 amino acids, between about 160 amino acids and about 250 amino acids, between about 180 amino acids and about 250 amino acids, between about 200 amino acids and about 250 amino acids, between about 220 amino acids and about 250 amino

acids, between about 10 amino acids and about 200 amino acids, between about 20 amino acids and about 200 amino acids, between about 40 amino acids and about 200 amino acids, between about 50 amino acids and about 200 amino acids, between about 60 amino acids and about 200 amino acids, between about 80 amino acids and about 200 amino acids, between about 90 amino acids and about 200 amino acids, between about 100 amino acids and about 200 amino acids, between about 120 amino acids and about 200 amino acids, between about 140 amino acids and about 200 amino acids, between about 150 amino acids and about 200 amino acids, between about 160 amino acids and about 200 amino acids, between about 180 amino acids and about 250 amino acids, between about 10 amino acids and about 150 amino acids, between about 20 amino acids and about 150 amino acids, between about 40 amino acids and about 150 amino acids, between about 50 amino acids and about 150 amino acids, between about 60 amino acids and about 150 amino acids, between about 80 amino acids and about 150 amino acids, between about 90 amino acids and about 150 amino acids, between about 100 amino acids and about 150 amino acids, between about 120 amino acids and about 150 amino acids, between about 140 amino acids and about 150 amino acids, between about 150 amino acids and about 150 amino acids, between about 160 amino acids and about 150 amino acids, between about 180 amino acids and about 250 amino acids, between about 160 amino acids and about 240 amino acids, about 161 amino acids, about 184 amino acids, and about 184 amino acids, within the N-terminal intrinsically disordered region of the RTE protein, or at the N-terminus of the RTE protein. In certain embodiments, the deletion is truncation from the N-terminal of the RTE protein. In certain embodiments, the RTE protein is a R2 protein. In certain embodiments, the deletion of the N-terminal IDR is relative to a WT R2-TG protein having an amino acid sequence set forth in SEQ ID NO: 1.

[0289] In certain embodiments, the engineered driver protein comprises a R2 RTE polypeptide comprising a deletion of a N-terminal intrinsically disordered region (IDR) or a portion thereof of the corresponding WT RTE protein. SEQ ID NOS: 679-714, 775-781 set forth amino acid sequences of exemplary driver proteins comprising a R2 RTE polypeptide comprising a deletion of a N-terminal intrinsically disordered region (IDR) or a portion thereof.

Exemplary engineered drivers

[0290] In certain embodiments, an engineered driver comprises one or more domains provided in the Examples. In certain embodiments, an engineered driver comprises or has an amino acid

sequence of any one of SEQ ID NOs 120-158, 160-171, 230-238, 250-267, 270-286, 618-757, 760-784, 790-1170, 1218-1327, 1336, 1337, 1368-1399, 1401-1410, 1417-1439, 1441-1715, 1725-1746, 1750-1873, and 1976-1989, or an amino acid sequence that is at least 70% identical (e.g., at least 75% identical, at least 80% identical, at least 85% identical, at least 90% identical, at least 91% identical, at least 92% identical, at least 93% identical, at least 94% identical, at least 95% identical, at least 96% identical, at least 97% identical, at least 98% identical, at least 99% identical, or 100% identical) thereto. In certain embodiments, an engineered driver is encoded by a DNA sequence (or a corresponding RNA version thereof) set forth in any one of SEQ ID NOs 10120-10158, 10160-10171, 10230-10238, 10250-10267, 10270-10286, 10618-10757, 10760-10784, 10790-101170, 101218-101327, 101336, 101337, 101368-101399, 101401-101410, 101417-101439, 101441-101715, 101725-101746, 101750-101873, and 101976-101989, or, a DNA sequence (or a corresponding RNA version thereof) that is at least 70% identical (e.g., at least 75% identical, at least 80% identical, at least 85% identical, at least 90% identical, at least 91% identical, at least 92% identical, at least 93% identical, at least 94% identical, at least 95% identical, at least 96% identical, at least 97% identical, at least 98% identical, at least 99% identical, or 100% identical) thereto. The above-noted amino acid and nucleotide sequences correspond to each other as follows: The SEQ ID NOS for the DNA sequences are systematically assigned to be greater by 10,000 compared to the SEQ ID NO for the amino acid sequence of the engineered driver that it encodes. As such, for example, the engineered driver having the sequence set forth in SEQ ID NO: 120 is encoded by a DNA having the sequence set forth in SEQ ID NO: 10120 (or by a corresponding RNA version thereof). In another example, the engineered driver having the sequence set forth in SEQ ID NO: 1976 is encoded by a DNA having the sequence set forth in SEQ ID NO: 11976 (or by a corresponding RNA version thereof).

[0291] In certain embodiments, an engineered driver has an amino acid sequence of SEQ ID NO 636 (encoded by SEQ ID NO: 10636). In certain embodiments, an engineered driver has an amino acid sequence of SEQ ID NO 618 (encoded by SEQ ID NO: 10618). In certain embodiments, an engineered driver has an amino acid sequence of SEQ ID NO 633 (encoded by SEQ ID NO: 10633). In certain embodiments, an engineered driver has an amino acid sequence of SEQ ID NO 1236 (encoded by SEQ ID NO: 11236). In certain embodiments, an engineered driver has an amino acid sequence of SEQ ID NO 1234 (encoded by SEQ ID NO: 11234). In certain embodiments, an engineered driver has an amino acid sequence of SEQ ID NO 668

(encoded by SEQ ID NO: 10668). In certain embodiments, an engineered driver has an amino acid sequence of SEQ ID NO 783 (encoded by SEQ ID NO: 10783). In certain embodiments, an engineered driver has an amino acid sequence of SEQ ID NO 3028 (encoded by SEQ ID NO: 13028). In certain embodiments, an engineered driver is at least 70% identical (e.g., at least 75% identical, at least 80% identical, at least 85% identical, at least 90% identical, at least 91% identical, at least 92% identical, at least 93% identical, at least 94% identical, at least 95% identical, at least 96% identical, at least 97% identical, at least 98% identical, at least 99% identical, or 100% identical) to any one of the aforementioned amino acid or DNA sequences.

[0292] The nucleic acids and polypeptides disclosed herein may be produced by methods known in the art. For example, the DNA or RNA molecules disclosed herein may be prepared synthetically via methods known in the art, or in the case of RNA, via in vitro transcription (IVT) methods known in the art. Likewise, the engineered driver and components thereof, e.g. RTE polypeptides, heterologous polypeptides, and other polypeptides and proteins described herein, may be produced via recombinant protein expression and purification, which is well suited for, e.g., fusion proteins. Methods for recombinant protein expression and purification are well known, and include those described by Green and Sambrook, *Molecular Cloning: A Laboratory Manual* (4th ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (2012)), the entire contents of which are incorporated herein by reference.

[0293] The terms “polynucleotide,” “nucleic acid” and “NA” may be used interchangeably and refer to a polymer of nucleotides. In certain embodiments, the polynucleotide comprises one or more chemical and/or sequence modifications. The polymer may include natural nucleosides (i.e., adenosine, thymidine, guanosine, cytidine, uridine, deoxyadenosine, deoxythymidine, deoxyguanosine and deoxycytidine), and nucleoside analogs having modified bases, modified sugars (e.g., 2'-fluororibose, 2'-methoxy), or modified phosphate groups (e.g., phosphorothioates, 2'-5' linkage). In certain embodiments, a nucleic acid comprises one or more chemical and/or sequence modifications. In certain embodiments, the modification is an RNA CAP, a modified polyA length (e.g., relative to a natural polyA), a chemically modified nucleotide, a 5' UTR (untranslated region) modification, a 3' UTR modification, a modified Sirloin (SINE-derived nuclear RNA localization) sequence, a modified (e.g., truncated) stem loop, an RNA stabilization motif (which may be a WPRE motif), a 5-methoxyuridine (5-MO-U) modification, a 5-methylcytidine (5mC) modification, or one or more additional and/or modified

microsatellites. In certain embodiments, a nucleic acid encoding an engineered driver and/or other gene (e.g., a therapeutic RNA or protein) is codon optimized (e.g., codon optimized for expression in human cells). In certain embodiments, codon optimization is implemented for RNA optimization. In certain embodiments, RNA optimization comprises one or more of the following modifications compared to the wildtype RNA molecule: In certain embodiments, RNA optimization comprises reducing the uracil (U) load of an RNA molecule. In certain embodiments, RNA optimization comprises reducing the GC% content of an RNA molecule. In certain embodiments, RNA optimization comprises reducing the length and/or number of intron sequences of an RNA molecule. In certain embodiments, RNA optimization comprises reducing RNA binding motifs or sites within an RNA molecule. In certain embodiments, RNA optimization comprises lowering ΔG (free energy) of an RNA molecule. In certain embodiments, RNA optimization comprises reducing the nucleotide repeats found in a sequence of an RNA molecule. In certain embodiments, RNA optimization comprises increased human or tissue tRNA frequency usage (AKA codon-usage) of an RNA molecule. In certain embodiments, RNA optimization comprises reducing the number of palindromic sequences in an RNA molecule. In certain embodiments, RNA optimization comprises maximizing pairing of bases of an in an RNA molecule. In certain embodiments, RNA optimization comprises removing splicing site sequences from an RNA molecule. In certain embodiments, RNA optimization comprises removing rare codons or other slowly translated codons. A person with skill in the art will be able to generate a polynucleotide encoding any one of the polypeptides disclosed herein, based on the polypeptide sequence, as provided herein.

[0294] In certain embodiments, an RNA driver construct encoding an engineered driver may include an RNA pseudoknot at the 5' and/or 3' end. In certain embodiments, the driver comprises an RNA pseudoknot. In certain embodiments, the gene delivery construct comprises an RNA pseudoknot. In certain embodiments, the pseudoknot is a EvopreQ1 pseudoknot (e.g., SEQ ID NO:70). In certain embodiments, the pseudoknot is a Mpknot1 pseudoknot (e.g., SEQ ID NO:71). In certain embodiments, the pseudoknot is a MALAT1 pseudoknot (e.g., SEQ ID NO:72).

[0295] In certain embodiments, and as will be described further herein below, a nucleic acid encoding an engineered driver may comprise a promoter. In certain embodiments, the promoter is a naturally occurring promoter. In certain embodiments, the promoter is a recombinant

promoter. In certain embodiments, the promoter is a constitutive, inducible, and/or tissue or cell specific promoter. The promoter may be selected from an EF1a promoter, a CMV promoter, an A1AT promoter, an Albumin gene promoter, an MNDu promoter, an MNDopt promoter, an SFFV promoter, or an ApoE promoter.

II. Retrotransposable-Element (RTE) Transgene Expression Systems

[0296] Retrotransposable elements, which may also be referred to herein as retroelements or RTEs, have naturally evolved to facilitate the incorporation of new DNA sequences into genomes. The RTE transgene expression systems of the disclosure take advantage of RTE function, and offer a method for introducing and integrating engineered transgenes of interest by guiding the synthesis of a DNA sequence from an RNA template, and permitting integration into host cell genomic DNA.

[0297] Generally, as provided herein, the RTE transgene expression systems of the disclosure comprise:

- (a) a driver construct comprising a nucleic acid encoding a driver comprising a RTE polypeptide; and
- (b) a gene delivery construct comprising:
 - (i) a nucleic acid encoding the transgene; and
 - (ii) at least one retroelement untranslated region (RTE-UTR),

wherein the at least one RTE-UTR is capable of binding the RTE polypeptide, and wherein the RTE polypeptide mediates integration of the nucleic acid encoding the transgene into a genome.

[0298] In certain embodiments, the driver encoded by the driver construct may be any one of the engineered drivers described elsewhere herein.

[0299] Descriptions and details of the components of each of the driver construct and gene delivery constructs of the RTE transgene expression systems of the disclosure are provided in subsequent sections.

[0300] In certain embodiments, the transgene comprises a promoter and an open reading frame encoding an RNA or protein. In certain embodiments, the transgene comprises stabilizing UTRs (e.g. 3' and/or 5' UTRs) that stabilize the RNA transcript of the transgene. In certain embodiments, the transgene includes a polyadenylation ("poly-A") signal. In certain embodiments, the transgene includes one or more introns.

[0301] It is noted that the driver construct may be DNA, RNA (e.g. mRNA), or a hybrid RNA/DNA molecule. Likewise, the gene delivery construct may be DNA or RNA (e.g. mRNA), or a hybrid RNA/DNA molecule. In exemplary embodiments, the driver construct and gene delivery construct are provided in trans, on separate DNA strands, separate RNA strands or separate RNA/DNA strands. In other certain embodiments, the driver construct and gene delivery construct are in a cis configuration and provided in a single nucleic acid. In certain embodiments, they are provided in cis as DNA. In certain embodiments, they are provided in cis as RNA. In certain embodiments, they are provided in cis as chimeric RNA/DNA.

[0302] For clarity, in a trans configuration, the gene delivery construct includes the transgene that may be flanked 5' and 3', independently, with a terminal region comprising one or more regulatory elements (e.g. RTE-UTRs, homology arms, etc.). In a cis configuration, the gene delivery nucleic acid portion includes an adjacent driver nucleic acid. The combined nucleic acid in cis configuration may be flanked 5' and 3', independently, with a terminal region comprising one or more regulatory elements (e.g. RTE-UTRs, homology arms, etc.).

[0303] In certain embodiments, the driver construct and the gene delivery construct can be provided as DNA, as RNA, or as a mixture of both. It is contemplated in other embodiments that the driver construct and the gene delivery construct can be provided as DNA, as RNA (e.g. as mRNA), or as a mixture of both encapsulated in a carrier, e.g. in a lipid nanoparticle (LNP). In yet other embodiments, the driver construct and the gene delivery construct can be provided in a delivery vector, e.g. a viral vector. These delivery formats are discussed in greater detail herein.

[0304] FIGS. 1A-1E schematically illustrate non-limiting examples of a gene delivery construct and/or a RTE transgene expression system.

[0305] FIG. 1A schematically illustrates a non-limiting example of a gene delivery construct comprising a transgene (e.g., a heterologous nucleic acid comprising a gene of interest) flanked

by two terminal regions, a 5' terminal region and a 3' terminal region, which may, independently, comprise a RTE-UTR, and optionally other elements.

[0306] A non-limiting example of a RTE transgene expression system in a trans configuration is illustrated in FIG. 1B. FIG. 1B schematically illustrates a gene delivery construct comprising a transgene flanked by two terminal regions, a 5' terminal region and a 3' terminal region as a first nucleic acid, and a driver construct comprising an engineered driver coding sequence encoding an engineered driver as a second nucleic acid that is distinct from the first nucleic acid. In some embodiments, the second nucleic acid does not contain terminal regions such that only the transgene in the gene delivery construct is integrated into the genome of a host cell, and the sequences encoding the engineered protein and any other genes that are on the second nucleic acid are not integrated into the host genome. The configuration illustrated in FIG. 1B shows only the engineered driver coding sequence encoded by the second nucleic acid. In some embodiments, one or more additional genes also may be included on the first and/or second nucleic acid. FIG. 1D shows the RTE transgene expression system in a trans configuration as illustrated in FIG. 1B, together with a genomic DNA before and after the transgene, along with the 3' and 5' terminal regions, has been integrated by the engineered driver expressed from the driver construct.

[0307] A non-limiting example of a RTE transgene expression system in a cis configuration is illustrated in FIG. 1C. FIG. 1C schematically illustrates a combined construct comprising (1) a driver construct comprising an engineered driver coding sequence encoding an engineered driver and (2) a gene delivery construct comprising a transgene, with the combined construct flanked by two terminal regions, a 5' terminal region and a 3' terminal region, FIG. 1C shows an embodiment in which the engineered driver coding sequence is upstream of the transgene. However, it will be appreciated that, in some embodiments, the transgene may be upstream of the engineered driver coding sequence. FIG. 1E shows the RTE transgene expression system in a cis configuration as illustrated in FIG. 1C, together with a genomic DNA before and after integration of the combined construct, along with the 3' and 5' terminal regions, by the engineered driver expressed from the driver construct.

[0308] It will be also appreciated that, in some embodiments, the engineered driver coding sequence is outside of the terminal regions, so that the only the transgene, and not the engineered driver coding sequence, is flanked by the 3' and 5' terminal regions. In such a case, the

transgene would be expected to be integrated into the genomic DNA, but not the driver coding sequence.

RTE-UTRs

[0309] As described elsewhere herein, a RTE transcript of a natural RTE comprises one or more untranslated regions (RTE-UTRs) that, as RNA, forms a structure that is capable of binding the RTE protein (e.g., at a compatible RNA binding domain comprised in the RTE protein), to facilitate the reverse transcription of the RTE transcript by the RTE protein. Whether in cis or trans configurations, the 5' terminal region and the 3' terminal region may comprise, independently, at least one RTE-UTR that is capable (e.g., as RNA) of binding a RTE polypeptide comprised in a driver encoded by the driver construct.

[0310] As noted elsewhere, the RTE protein and the one or more RTE-UTRs of a given RTE typically have co-evolved to have sufficient binding affinity for each other. As such, in certain embodiments, the one or more RTE-UTRs employed in the RTE transgene expression system may be derived from the same RTE as the RTE polypeptide comprised in the driver. However, it has also been found that certain heterologous pairs RTE-UTRs and RTE polypeptides, where the RTE-UTRs and the RTE polypeptides are derived from different RTEs (e.g., different species of R2 RTEs) have sufficient affinity to facilitate interaction of the gene delivery construct and the driver. Some heterologous pairings may be even more effective in facilitating transgene integration than homologous pairing. For example, as shown in Example 10, in a system comprising a R2-1_TG driver construct encoding an engineered driver comprising a R2-1_TG RTE polypeptide, a reporter construct (a R2-1-PTT reporter) comprising a 5' RTE-UTR (SEQ ID NO: 35027) and a 3' RTE-UTR (SEQ ID NO: 33027) from R2-1-Phylloscopus trochilus (PTT), was more effective in facilitating transgene integration than another reporter construct comprising a RTE-UTR derived from R2-1-TG. As such, in certain embodiments, the one or more RTE-UTRs employed in the RTE transgene expression system may be derived from a different RTE as the RTE polypeptide comprised in the driver.

[0311] A RTE-UTR may be referred to by the species of RTE from which it is derived. For example, a RTE-UTR derived from R2-1-TG may be referred to herein as a "R2-1-TG RTE-UTR". Similarly, a gene expression construct may be referred to herein by the RTE from which the one or more RTE-UTRs comprised therein are derived. As such, a gene delivery construct

comprising one or more RTE-UTRs derived from the R2-1-TG RTE may be referred to herein as a “R2-1-TG gene delivery construct” or “R2-1-TG reporter”.

[0312] A RTE-UTR comprised in a 5' terminal region of the gene delivery construct (when in trans configuration) or of the combined construct (when in cis configuration) may be referred to herein as a 5' RTE-UTR. A RTE-UTR comprised in a 3' terminal region of the gene delivery construct (when in trans configuration) or of the combined construct (when in cis configuration) may be referred to herein as a 3' RTE-UTR. In certain embodiments, RTE-UTRs comprised in the gene delivery construct (when in trans configuration) or the combined construct (when in cis configuration) may be RTE-UTRs derived from natural RTEs (e.g. from any one of the RTEs, including R2 RTEs, described herein). In certain embodiments, the RTE-UTRs may be modified RTE-UTRs based on RTE-UTRs from one or more natural RTEs.

[0313] In some embodiments, the 3' RTE-UTR, e.g., of a R2-TG, may comprise a conserved Stem Loop (SL) region and a variable number of a microsatellite repeats (e.g., a minimal 3' UTR required for efficient transgene integration).

[0314] In certain embodiments, the one or more RTE-UTR(s) are derived from a non-LTR RTE. In certain embodiments, the one or more RTE-UTR(s) are derived from a Class I transposable element that utilizes an intermediate RNA during integration by a copy-and-paste mechanism (PMIDs: 32955944 and 18261821) and that lacks Long Terminal Repeats (LTR) or Inverted Terminal Repeats (ITR) in their structure, and that can be derived from a superclade selected from the group consisting of Retrons (include clades AbiA, AbiK, AbiP2, CRISPR, CRISPR-like, DCRs, G2L4, G2L, G2Lb, G2Lc, GII, Retrons, UG1, UG2, UG3, UG4, UG5, UG6, UG7, UG8, UG9, UG10, UG11, UG12, UG13, UG14, UG15, UG16, UG17, UG18, UG19, UG20, UG21, UG22, UG23, UG24, UG25, UG26, UG27, UG28 and UG28b), Group II introns (include clades E, ML, B, CL1A, CL1B, CL2A, CL2B, F, D and C), Penelope-like retroelements (include clades Pen1, Pen1a, Pen2, Pen3, Pen3a, Pen3b, Pen4, Pen5, Pen5a and Pen6), and non-LTR retroelements (include clades CRE, R4, Hero, NeSL, R2, RandI, Proto1, L1, Tx1, RTEPT, Proto2, RTEX, RTE, Outcast, Ingi (which includes sub-clade Vingi), I, Nimb, Tad1, Loa, R1, Jockey, Rex1, CR1, L2, L2A, L2B, Daphne, and Crack).

[0315] In certain embodiments, a RTE transgene expression systems of the disclosure may utilize a RTE-UTR (a 3' RTE-UTR and/or a 5' RTE-UTR) derived from a R2 RTE.

[0316] Non limiting examples of RE-UTRs from R2 retroelements found in a variety of organisms are set forth in SEQ ID NOS: 215-220, 20500-20712 and 20754-20755. Of these, 5'RTE-UTRs are set forth in SEQ ID NOS: 215, 217, 219, 20500-20605, and 20754 and 3' RTE-UTRs are set forth in SEQ ID NOS: 216, 218, 220, 20606-20712, and 20755. Other non-limiting examples of 5' RTE-UTRs are set forth in SEQ ID NOS: 20715-20723. Other non-limiting examples of 3'RTE-UTRs are set forth in SEQ ID NOS: 20713-20714 and 20724-20728. The RTE-UTR sequences in the present application are provided as DNA sequences. For each RTE-UTR sequence provided, it will be understood that the present disclosure also includes a respective RNA version thereof.

[0317] In certain embodiments, a RTE transgene expression systems of the disclosure may utilize a RTE-UTR (a 3' RTE-UTR and/or a 5' RTE-UTR) derived from one of the following R2 RTEs, as listed in Table 3:

Table 3: RTE-UTRs of representative R2 RTEs

	RTE name	Organism species	SEQ ID NO for 5' RTE- UTR	SEQ ID NO for 3' RTE- UTR
1	R2-1-TG	Taeniopygia guttata	215	216
2	R2-ZA	Zonotrichia albicollis	217	218
3	R4-AC	Anolis carolinensis	219	220
4	R4-1-PH	Parhyale hawaiiensis	20754	20755

Table 3 provides the DNA sequence of the 5' RTE-UTR and the 3' RTE-UTR from each of the listed RTE species. The RTE-UTR sequences in Table 3 are provided as DNA sequences. For each RTE-UTR sequence provided, it will be understood that the present disclosure also includes a respective RNA version thereof.

[0318] In certain embodiments, a RTE transgene expression systems of the disclosure may utilize a RTE-UTR (a 3' RTE-UTR and/or a 5' RTE-UTR) derived from one of the following R2 RTEs, as listed in Table 4:

Table 4: RTE-UTRs of R2 RTEs

	RTE name	Abbreviation	Organism species	SEQ ID NO for 5' RTE-UTR	SEQ ID NO for 3' RTE-UTR
1	R2-1_Acridotheres_tristis	R2-AT	Acridotheres tristis	35001	33001
2	R2-1_Acrocephalus_scorpaeus	R2-ASS	Acrocephalus scorpaeus	35002	33002
3	R2-1_Agelaius_tricolor	R2-AgT	Agelaius tricolor	35003	33003
4	R2-1_Ammodramus_caudacutus	R2-AC	Ammodramus caudacutus	35004	33004
5	R2-1_Ammospiza_nelsoni	R2-AN	Ammospiza nelsoni	35005	33005
6	R2-1_Buteo_platypterus	R2-BP	Buteo platypterus	35006	33006
7	R2-1_Chamaeotylas_choloensis	R2-CC	Chamaeotylas choloensis	35007	33007
8	R2-1_Chamaeotylas_fuelleborni	R2-CF	Chamaeotylas fueleborni	35008	33008
9	R2-1_Cichladusa_arquata	R2-CA	Cichladusa arquata	35009	33009
10	R2-1_Ciconia_episcopus	R2-CE	Ciconia episcopus	35010	33010
11	R2-1_Colaptes_auratus	R2-CoAu	Colaptes auratus	35011	33011
12	R2-1_Cossypha_archeri	R2-CoAr	Cossypha archeri	35012	33012
13	R2-1_Cossypha_dichroa	R2-CD	Cossypha dichroa	35013	33013
14	R2-1_Falco_fasciinuca	R2-FF	Falco fasciinuca	35014	33014
15	R2-1_Gallinago_paraguaiae	R2-GP	Gallinago paraguaiae	35015	33015
16	R2-1_Garrulus_glandarius	R2-GG	Garrulus glandarius	35016	33016
17	R2-1_Geranoaetus_albicaudatus	R2-GA	Geranoaetus albicaudatus	35017	33017
18	R2-1_Harpyopsis_novaeguineae	R2-HN	Harpyopsis novaeguineae	35018	33018
19	R2-1_Limnodromus_scolopaceus	R2-LS	Limnodromus scolopaceus	35019	33019
20	R2-1_Melierax_canorus	R2-MC	Melierax canorus	35020	33020
21	R2-1_Myiozetetes_cayanensis	R2-MyC	Myiozetetes cayanensis	35021	33021
22	R2-1_Otis_tarda	R2-OT	Otis tarda	35022	33022

	RTE name	Abbreviation	Organism species	SEQ ID NO for 5' RTE-UTR	SEQ ID NO for 3' RTE-UTR
23	R2-1_Prinia_subflava	R2-PS	Prinia subflava	35023	33023
24	R2-1_Serinus_canaria	R2-SC	Serinus canaria	35024	33024
25	R2-1_Aphelocoma_californica	R2-ApC-1	Aphelocoma californica	35025	33025
26	R2-1_Oenanthe_melanoleuca	R2-OM	Oenanthe melanoleuca	35026	33026
27	R2-1_Phyloscopus_trochilus_trochilus	R2-PTT	Phylloscopus trochilus trochilus	35027	33027
28	R2-1_Geothlypis_trichas	R2-GT	Geothlypis trichas	35028	33028
29	R2-1_Xenus_cinereus	R2-XC	Xenus cinereus	35029	33029
30	R2-2_Aphelocoma_californica	R2-2-ApC	Aphelocoma californica	35030	33030
31	R2-2_Cichladusa_arquata	R2-CiA	Cichladusa arquata	35031	33031
32	R2-1_Scolopax_minor	R2-SM	Scolopax minor	35032	33032
33	R2-1_Tyto_alba	R2-Ta	Tyto alba	35033	33033
34	R2-2_Acrocera_orbiculus	R2-AcC	Acrocera orbiculus	35034	33034
35	R2-1_Drosophila_sulfurigaster_sulfurigaster	R2-DSS	Drosophila sulfurigaster sulfurigaster	35035	33035
36	R2-1_Rhopalosiphum_padi	R2-RP	Rhopalosiphum padi	35036	33036
37	R2-1_Nemotelus_nigrinus	R2-NN	Nemotelus nigrinus	35037	33037
38	R2-1_Chlorops_oryzae	R2-CO	Chlorops oryzae	35038	33038
39	R2-2_Andrena_bucephala	R2-AB	Andrena bucephala	35039	33039
40	R2-3_Ascidia_mentula	R2-AM	Ascidia mentula	35040	33040
41	R2-1_Crotalus_oreganus_helleri	R2-COH	Crotalus oreganus helleri	35041	33041
42	R2-1_Gopherus_flavomarginatus	R2-GF	Gopherus flavomarginatus	35042	33042
43	R2-1_Malaclemys_terrapi_n_pileata	R2-MTP	Malaclemys terrapi_n pileata	35043	33043
44	R2-1_Eublepharis_macularius	R2-EM	Eublepharis macularius	35044	33044
45	R2-1_Emys_orbicularis	R2-EO	Emys orbicularis	35045	33045
46	R2-1_Caretta_caretta	R2-CC	Caretta caretta	35046	33046
47	R2-1_Pleurodeles_waltl	R2-PW	Pleurodeles waltl	35047	33047

	RTE name	Abbreviation	Organism species	SEQ ID NO for 5' RTE-UTR	SEQ ID NO for 3' RTE-UTR
48	R2-1_Aspidoscelis_tigris_stejnegeri	R2-1-ATS	Aspidoscelis tigris stejnegeri	35048	33048
49	R2-1_Aspidoscelis_tigris_stejnegeri	R2-2-ATS	Aspidoscelis tigris stejnegeri	35049	33049
50	R2-1_Mauremys_mutica	R2-1-MM	Mauremys mutica	35050	33050
51	R2-3_Mauremys_mutica	R2-1-MM	Mauremys mutica	35051	33051
52	R2-1_Xiphophorus_hellerii	R2-XH	Xiphophorus hellerii	35052	33052
53	R2-1_Gambusia_affinis	R2-GamA	Gambusia affinis	35053	33053
54	R2-1_Schizopygopsis_malacanthus	R2-SMa	Schizopygopsis malacanthus	35054	33054
55	R2-1_Schizopygopsis_pylzovi	R2-1-SP	Schizopygopsis pylzovi	35055	33055
56	R2-2_Schizopygopsis_pylzovi	R2-2-SP	Schizopygopsis pylzovi	35056	33056
57	R2-1_Gila_orcuttii	R2-GO	Gila orcuttii	35057	33057
58	R2-1_Ahaetulla_prasina	R2-AP	Ahaetulla prasina	35058	33058
59	R2-1_Aldabrachelys_gigantea	R2-AG	Aldabrachelys gigantea	35059	33059
60	R2-1_Ascidia_mentula	R2-AsMe	Ascidia mentula	35060	33060

[0319] Table 4 provides the DNA sequence of the 5' RTE-UTR and the 3' RTE-UTR from each of the listed RTE species. The RTE-UTR sequences in Table 4 are provided as DNA sequences. For each RTE-UTR sequence provided, it will be understood that the present disclosure also includes a respective RNA version thereof.

[0320] In certain embodiments, a R2 RTE-UTR used in a RTE transgene integration system (e.g., comprised in a reporter construct) may have an identical sequence with the RTE-UTR sequence comprised in a natural R2 RTE, for example as listed in the tables above. In certain embodiments, the R2 RTE-UTR used in a RTE transgene integration system (e.g., comprised in a reporter construct) may have least 70%, at least 71%, at least 72%, at least 73%, at least 74%, at least 75%, at least 76%, at least 77%, at least 78%, at least 79%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at

least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% sequence identity with the RTE-UTR sequence comprised in a natural R2 RTE, for example as listed in the tables above.

[0321] An “ancestral reconstruction process” refers to the method of extrapolating back in time and to model the ancient parents of the identified sequences, (see for example Campitelli et al., Reconstruction of full-length LINE-1 progenitors from ancestral genomes. Genetics. 2022 Jul; 221(3): iyac074.). In certain embodiments, an R2 RTE-UTR used in a RTE transgene integration system (e.g., comprised in a reporter construct) may be an “ancestral RTE-UTR” based on an ancestral reconstruction based on a plurality of R2 RTE species.

[0322] In certain embodiments, a RTE transgene expression systems of the disclosure may utilize an ancestral RTE-UTR (a 3’ RTE-UTR and/or a 5’ RTE-UTR) derived from one of the following combinations of R2 RTEs, as listed in Table 5 (Examples 11 and 25 provide the reconstruction method):

Table 5: Ancestral RTE-UTRs

	RTEs included in ancestral reconstruction process	SEQ ID NO for 5' RTE-UTR	SEQ ID NO for 3' RTE-UTR
1	3utr-AR1-4 (SEQ ID NO:1338, 1339 and 1340)	2886	3003
2	3utr-AR2-5(SEQ ID NO:1344, 1345, 1346 and 1347)	2886	3004
3	3utr-AR3-7(SEQ ID NO:1352, 1353, 1354)	2886	3005
4	5utr-AR1-4(SEQ ID NO: 1441, 1442, 1443,)	1845	1866
5	5utr-AR2-5(SEQ ID NO:1352, 1353, 1354)	1874	1866
6	5utr-AR3-4 (SEQ ID NO:1355, 1356, 1357)	3006	1866
7	5utr-AR4-6 (SEQ ID NO:1358, 1359)	3007	1866
8	5utr-AR5-4 SEQ ID NO: 1360, 1361 and 1362)	3008	1866

	RTEs included in ancestral reconstruction process	SEQ ID NO for 5' RTE-UTR	SEQ ID NO for 3' RTE-UTR
9	5utr-AR7-5 (SEQ ID NO:1363, 1364, 1365)	1846	1866
10	5utr-AR8m-10 (SEQ ID NO:1356, 1357)	1847	1866
11	5utr-AR8-11 (SEQ ID NO:1355, 1358, 1359)	2883	1866
12	5utr-AR8f-10 (SEQ ID NO:1356, 1366)	1848	1866

[0323] The RTE-UTR sequences in Table 5 are provided as DNA sequences. For each RTE-UTR sequence provided, it will be understood that the present disclosure also includes a respective RNA version thereof.

Homology arms

[0324] In certain embodiments, the 5' terminal region and the 3' terminal region (e.g., comprised at the 5' and 3' ends of a gene delivery construct) may comprise, independently, a homology arm with nucleic acid sequences having homology to nucleic acid sequences in the human genome. In some embodiments, each homology arm is independently selected and is from about 4 to about 200 nucleotides in length or more, for example about 4, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 200 or more nucleotides long. In some embodiments, the homology arm correspond to a sequence in the 28S rDNA locus in the human genome. In some embodiments, the homology arm correspond to a sequence in the AAVS1 locus in the human genome. In some embodiments, the nucleic acid sequence of the homology arm is in a reading frame that is different than the open reading frame of the transgene comprised in the gene delivery construct. In some embodiments, the nucleic acid sequence of the homology arm is in the same reading frame as the transgene.

[0325] In some embodiments, a gene delivery construct comprises (e.g. in a 5' terminal region and/or a 3' terminal region) one or more sequences (e.g. homology arms) that are homologous to a target sequence (e.g., a target sequence in a host genome). Non-limiting examples of target sequences include safe harbor genomic targets. In some non-limiting embodiments, a safe harbor

genomic target is a AAVS1, a 28SrDNA, a hROSA26, a CCR5, a SHS231, and a PCSK9 safe harbor locus.

Other optional gene delivery construct elements

[0326] In some embodiments, a gene delivery construct may comprise one or more RNA nuclear localization sequences (e.g., a SAFB motif) and/or one or more stabilization motifs (e.g., a WPRE motif). In some embodiments, the gene delivery construct also may comprise flanking regions homologous to a target sequence in a genome.

[0327] In certain embodiments, the gene delivery construct nucleic acid comprises one or more chemical or sequence modifications. In certain embodiments, the modification is a RNA CAP, a modified polyA length (e.g., relative to a natural polyA), a chemical modification (e.g., a pseudouridine and/or a methylpseudouridine), a 5' UTR modification, a 3' UTR modification, a modified Kozak sequence, a modified (e.g., truncated) stem loop, a RNA stabilization motif (which may be a WPRE motif), a 5-methoxyuridine (5-MO-U) modification, a 5-methylcytidine (5mC) modification, or one or more additional and/or modified microsatellites. In certain embodiments, a nucleic acid encoding an engineered driver and/or a transgene (e.g., encoding a therapeutic RNA or protein) is codon optimized (e.g., codon optimized for expression in human cells). In certain embodiments, codon optimization is for RNA optimization. In certain embodiments, RNA optimization comprises reducing the Uracil (U) load of a RNA molecule.

Transgene

[0328] A gene delivery construct of the disclosure comprises a sequence of interest, which may be referred to as a transgene, which encodes a RNA or protein (e.g., a therapeutic RNA or protein), destined for integration into a genome. In certain embodiments, a gene delivery construct may comprise a plurality of transgenes.

[0329] As noted above, the gene delivery construct may comprise (e.g. in a 5' terminal region and/or a 3' terminal region) one or more sequence elements (e.g., RTE-UTRs) that interact with the RTE polypeptide comprised in the driver, e.g., that is encoded in the driver construct).

[0330] In certain embodiments, a gene delivery construct comprises a promoter, an open reading frame encoding a RNA or protein (the transgene), and optionally a polyA sequence.

[0331] In some embodiments, the promoter may be an inducible promoter. In some embodiments, the promoter is a tissue-specific promoter. In certain embodiments, the promoter may be, e.g., a MNDopt promoter, a MNDu promoter, a EF1a promoter, a CMV promoter, a A1AT promoter, a Albumin promoter, or a ApoE promoter.

[0332] In certain embodiments, a gene delivery construct may comprise a segment (e.g. a transgene) encoding a detection polypeptide, e.g. useful to detect integration of the transgene by a given driver. In certain embodiments, a detection polypeptide is a human influenza hemagglutinin (HA) Flag (DYKDDDDK; SEQ ID NO: 30), green fluorescent protein (GFP) or variants such as EGFP), or mCherry.

[0333] In some embodiments, the transgene encodes an RNA of interest, e.g., a therapeutic RNA, a regulatory RNA, or an RNA enzyme. In some embodiments, the therapeutic RNA is a messenger RNA (mRNA), antisense RNA (asRNA), RNA interference (RNAi), or an RNA aptamer. In some embodiments, the therapeutic RNA is an mRNA that encodes a therapeutic protein.

[0334] In some embodiments, a gene of interest encodes a therapeutic RNA, a regulatory RNA, an mRNA, an RNA enzyme, or other RNA. A regulatory RNA can be a siRNA, a miRNA, an antisense RNA, or other regulatory RNA. In some embodiments, an encoded RNA is an aptamer. An mRNA can be an RNA that encodes a protein of interest (for example a protein that has therapeutic, diagnostic, and/or other properties). Non limiting examples of proteins of interest include antibodies, regulatory proteins, hormones, cytokines, structural proteins, enzymes, membrane proteins, and other useful therapeutic or diagnostic proteins. Such proteins can be naturally occurring proteins or modified proteins (e.g., containing one or more amino acid substitutions relative to a naturally occurring counterpart protein). In some embodiments, a heterologous nucleic acid can encode one or more genes of interest (e.g., one or more regulatory RNAs and/or proteins of interest).

[0335] In some embodiments, non-limiting examples of a gene of interest include the genes that encode Factor VIII, Factor IX, Phenylalanine hydroxylase, ATP7B, alpha glucosidase, argininosuccinate synthetase, galactose-1-phosphate uridylyltransferase, ornithine transcarbamylase, or the like. A person with skill in the art will select the gene(s) of interest to be delivered to a subject's cell and inserted into a subject's genome.

[0336] In some embodiments, the heterologous nucleic acid encodes two or more genes of interest.

[0337] In certain embodiments, the transgene comprises stabilizing UTRs (e.g. 3' and/or 5' UTRS) that stabilize the RNA transcript of the transgene. In certain embodiments, the transgene includes one or more introns.

ORF1-like RNA-Binding Protein (ORBP) system

[0338] In some aspects, provided is a gene delivery and insertion system in which an RNA (or in some embodiments DNA) molecule containing the transgene (i.e., gene delivery construct) further includes an expression cassette for a cis-acting RNA-binding activity. In some embodiments, the expression cassette comprises a coding sequence for a cis-acting RNA-binding protein. The cis-acting RNA binding protein is translated in cis with the transgene, thus, allowing proximity between the RNA binding protein and the transgene. FIG. 3 depicts three non-limiting examples of such configurations. Without wishing to be bound to theory, the extended local level of the RNA binding protein is expected to support efficient binding to the transgene. The binding of the RNA binding protein to the transgene is expected to both protect the transgene from nucleases and immune sensor proteins. Protection from nucleases can result in more stable and long-lasting transgene. The competition with immune sensor protein can protect from immune activation in the cells. The RNA binding protein should not interfere with the RT binding. In some embodiments, the cis-acting RNA-binding protein is derived from the ORF1p of a retrotransposable element. The RBP may be further engineered, with amino acid substitutions deletions and the like. The ORF1p may be derived from any vertebrate or non-vertebrate organism. In some embodiments, the ORF1p is from Mouse LINE-1 retrotransposable element. In some embodiments, the ORF1p is from ZFL2-1 retrotransposable element from Zebrafish. In some embodiments, the ORF1p is from DOC2 retrotransposable element from Drosophila. In some embodiments, the ORF1p is from Maui retrotransposable element from Fugu fish. In some embodiments, the ORF1p is from I-3 retrotransposable element from Zebrafish. Non-limiting examples of ORBP constructs are set forth in SEQ ID NOs: 195-204 and Example 7. The ORF1p may be further engineered, with amino acid substitutions deletions and the like. In some embodiments, the ORF1-like RNA binding protein is positioned upstream of the transgene, i.e., at the 5' end of the gene delivery RNA. In some embodiments an RNA element that inhibits RT is positioned between the RNA binding protein ORF and the

transgene to avoid reverse transcription of the RNA binding protein. In some embodiments, the coding of the RNA binding protein is selected to ensure maximum stop codons at the end of its coding frame, so that upon readthrough of the transgene the RNA binding protein will not support translation. In some embodiments, more than one stop codon is placed at the end of the transgene to avoid readthrough from the transgene to the RNA binding protein. In some embodiments, two or three or more stop codons are placed at the end of the transgene. Non-limiting examples of polypeptides, which include stop codons are set forth in SEQ ID NOs: 195-199.

III. *Delivery systems*

[0339] The present disclosure also describes delivery systems for use in delivery of the RTE transgene expression systems (also referred to herein as “RTE system”) of the disclosure to a recipient host cell, and promoting integration of the transgene into a target nucleic acid (e.g., a genomic nucleic acid) in the host cell, either in vitro, in vivo, or ex vivo.

[0340] In certain embodiments, the delivery systems of the disclosure are useful for high efficiency integration of a transgene into a target region (e.g., a sequence specific target) within a target nucleic acid (e.g., within a host genome). In certain embodiments, the delivery systems of the disclosure are useful to redirect integration of a transgene to a particular target region (e.g., relative to integration mediated by a naturally occurring protein).

[0341] In certain embodiments, the RTE system is delivered using nanoparticles (of any type). In certain embodiments, the RTE system is delivered using lipid nanoparticles (LNPs). LNPs are particles composed of lipids, typically spherical in shape and a micron or less in diameter, that are used as delivery vehicles for genetic material, such as DNA or RNA, into cells. Without being limited by theory, LNPs protect the genetic material from degradation and facilitate its entry into target cells.

[0342] In certain embodiments, the average size of the LNP is between about 10nm and about 1000nm in diameter, between about 100nm and about 1000nm in diameter, between about 100nm and about 1000nm in diameter, between about 500nm and about 1000nm in diameter, between about 10nm and about 100nm in diameter, between about 10nm and about 200nm in diameter, between about 10nm and about 500nm in diameter, or between about 100nm and about

500nm in diameter. Any technique known in the art may be used to determine the size of the LNP. For example, LNP size could be measured using dynamic light scattering (DLS).

[0343] In certain embodiments, the driver and gene delivery constructs are provided as RNAs and packaged into the LNP. In certain embodiments, a driver construct is provided as RNA in a different LNP as the gene delivery construct RNA. In certain embodiments, a LNP that comprises RNAs encoding a gene delivery construct, comprise a driver protein.

[0344] Generally, LNPs consists of four components: ionizable cationic lipids, phospholipids, cholesterol, and PEGylated lipids. In certain embodiments, the LNP comprised an ionizable lipid, a PEGylated lipid, an amino lipid, a phospholipid, a cholesterol, a sterol, a non-cationic lipid, or any combination thereof. In some embodiments, the ionizable lipid is an ionizable cationic lipid. In certain embodiments, the LNP of the disclosure comprises: an ionizable cationic lipid, a phospholipid, a cholesterol, and a PEGylated lipid.

[0345] In certain embodiments, the RTE system is delivered using a targeted LNP (also referred to herein as “tLNP”). In certain embodiments, the tLNP is coated with an antibody or antigen binding domain that recognizes a target antigen, which can also increase specificity of delivery.

[0346] In other embodiments, the RTE system is delivered using a viral vector-based delivery system. Contemplated viral vector delivery systems include, but are not limited to, adenovirus-, AAV-, and lentiviral-based viral vector systems. In certain embodiments, the nucleic acid constructs further comprise additional flanking nucleic acids. In certain embodiments, the additional flanking nucleic acids are Adeno-associated viruses (AAV) or lentiviral nucleic acids. In certain embodiments, the additional flanking nucleic acids are AAV inverted terminal repeats (ITRs).

IV. Pharmaceutical Compositions

[0347] Also provided herein are pharmaceutical compositions comprising any of the nucleic acid constructs, RTE transgene expression systems, and/or delivery systems of the disclosure, and optionally a pharmaceutically acceptable carrier, adjuvant, diluent, or excipient.

V. Methods of Generating Cells

[0348] The methods, systems, and compositions described herein are useful for delivering a transgene to a host cell and promoting integration of the transgene into a target nucleic acid (e.g., in the genome of the cell).

[0349] Accordingly, in certain embodiments, provided herein are methods of promoting the integration of a transgene into a genome of a cell *ex vivo*, the method comprising contacting a population of cells with (a) a RTE transgene expression system described herein, (b) a nanoparticle or LNP comprising a RTE transgene expression systems described herein, (c) a viral vector encoding the components of the RTE transgene expression systems described herein, or (c) or a pharmaceutical composition comprising any one of the same. In these embodiments, the contacting takes place *ex vivo*, e.g. with the cells being prepared for a cell therapy. The cells may be either allogeneic or autologous for the intended recipient.

[0350] Likewise, in certain embodiments, provided herein are methods of promoting the integration of a transgene into a genome of a cell *in vivo*, the method comprising administering to a subject (a) a RTE transgene expression system described herein, (b) a nanoparticle or LNP comprising a RTE transgene expression systems described herein, (c) a viral vector encoding the components of the RTE transgene expression systems described herein, or (c) or a pharmaceutical composition comprising any one of the same.

[0351] The host cell can be any type of mammalian cell. In certain embodiments, the host cell is a quiescent cell. In certain embodiments, the host cell is an actively dividing cell.

VI. Methods of Treatment

[0352] The methods, systems, and compositions described herein are useful for delivering a transgene of the disclosure to a subject in need of treatment. Accordingly, In certain embodiments, provided herein are methods of treating a disease in a subject in need thereof, comprising administering to the subject a therapeutically effective amount of: (a) RTE transgene expression systems described herein, (b) a nanoparticle or LNP comprising a RTE transgene expression systems described herein, (c) a viral vector encoding the components of the RTE transgene expression systems described herein, or (d) or a pharmaceutical composition comprising any one of the same.

[0353] In certain embodiments, the subject is human.

[0354] In certain embodiments, the route of administration includes, but is not limited to intravenous, intraperitoneal, intramuscular, subcutaneous, intrathecal, and intradermal administration.

[0355] There is also provided herein transgene expression systems (as well as associated LNPs and pharmaceutical compositions) for use in a method of treating the human or animal body by therapy.

[0356] There is also provided herein a use of the transgene expression systems (as well as associated LNPs and pharmaceutical compositions) as described herein for the manufacture of a medicament.

Enumerated Embodiments

Set I:

[0357] Some embodiments of this disclosure relate to Embodiment set I, as follows:

[0358] Embodiment I-1. A nucleic acid encoding an engineered protein comprising a retroelement-derived polypeptide fused to at least one heterologous polypeptide,

wherein the retroelement-derived polypeptide comprises an amino acid sequence derived from an R2 group retroelement;

wherein the at least one heterologous polypeptide comprises an RNA/DNA processing polypeptide, an RNA/DNA repair polypeptide, a nucleic acid binding polypeptide, a nucleosome binding polypeptide, or any combination thereof; and

wherein the engineered protein exhibits at least one of efficiency, specificity, accuracy, fidelity or processivity compared to reference retroelement polypeptide.

[0359] Embodiment I-2. The nucleic acid of embodiment I-1, wherein the retroelement-derived polypeptide comprises a reverse transcriptase domain, an endonuclease domain, and/or an RNA binding domain.

[0360] Embodiment I-3. The nucleic acid of embodiment I-1 or I-2, wherein the retroelement-derived polypeptide comprises one or more of an amino acid substitution, amino acid deletion or amino acid truncation, when compared to the reference retroelement polypeptide.

[0361] Embodiment I-4. The nucleic acid of any one of embodiments I-1 to I-3, wherein the R2 group retroelement comprises a R2-TG derived retroelement, R2-ZA derived retroelement, R4-AC derived retroelement, R4-1_PH derived retroelement, R2-1-AgTr derived retroelement, R2-1-ChCh derived retroelement, R2-1-CiAr derived retroelement, R2-1-CiEp derived retroelement, R2-1-CiSt derived retroelement, R2-1-ErRu derived retroelement, R2-1-GaPa derived retroelement, R2-1-GeAl derived retroelement, R2-1-HaNo derived retroelement, R2-1-LiSc derived retroelement, R2-1-MeCa derived retroelement, R2-1-OeMe derived retroelement, R2-1-OtTa derived retroelement, R2-1-TyAl derived retroelement, R2-1-XeCi derived retroelement, R2-1-ApCa derived retroelement, R2-1-PhTr derived retroelement, R2-1-PsHy derived retroelement, R2-1-AnBu derived retroelement, R2-1-BrLa derived retroelement, R2-1-AcOr derived retroelement, R2-2-AcOr derived retroelement, R2-2-BrLa derived retroelement, R2-2-AnBu derived retroelement, R2-3-AcOr derived retroelement, R2-4-AcOr derived retroelement, R2-5-AcOr derived retroelement, R2-1-MeAu derived retroelement, R2-2-MeAu derived retroelement, R2-1-DrSu derived retroelement, R2-1-RhPa derived retroelement, R2-6-AcOr derived retroelement, R2-1-NeNi derived retroelement, R2-7-AcOr derived retroelement, R2-1-ChOr derived retroelement, R2-3-AnBu derived retroelement, R2-1-AcTr derived retroelement, R2-1-AcSc derived retroelement, R2-2-AgTr derived retroelement, R2-1-AmCa derived retroelement, R2-1-AmNe derived retroelement, R2-1-AqAf derived retroelement, R2-1-BuPl derived retroelement, R2-2-ChCh derived retroelement, R2-1-ChFu derived retroelement, R2-2-CiAr derived retroelement, R2-2-CiEp derived retroelement, R2-2-CiSt derived retroelement, R2-1-CoAu derived retroelement, R2-1-CoAr derived retroelement, R2-1-CoDi derived retroelement, R2-1-FaFa derived retroelement, R2-2-GaPa derived retroelement, R2-1-GaGl derived retroelement, R2-2-GeAl derived retroelement, R2-2-HaNo derived retroelement, R2-2-LiSc derived retroelement, R2-2-MeCa derived retroelement, R2-1-MyCa derived retroelement, R2-2-OtTa derived retroelement, R2-1-PrSu derived retroelement, R2-1-SeMi derived retroelement, R2-1-SeCa derived retroelement, R2-1-SeKi derived retroelement, R2-2-TyAl derived retroelement, R2-2-ApCa derived retroelement, R2-2-ErRu derived retroelement, R2-1-GeTr derived retroelement, R2-1-LiCa derived retroelement, R2-2-OeMe derived retroelement, R2-2-PhTr derived retroelement, R2-1-MeGe derived retroelement, R2-2-GeTr derived

retroelement, R2-2-XeCi derived retroelement, R2-3-ApCa derived retroelement, R2-3-CiAr derived retroelement, R2-3-ErRu derived retroelement, R2-2-MyCa derived retroelement, R2-3-PhTr derived retroelement, or R2-2-ScMi derived retroelement.

[0362] Embodiment I-5. The nucleic acid of embodiment I-4, wherein the R2 group retroelement comprises a R2-TG derived retroelement.

[0363] Embodiment I-6. The nucleic acid of embodiment I-4, wherein the R2 group retroelement comprises a R2-ZA derived retroelement.

[0364] Embodiment I-7. The nucleic acid of embodiment I-4, wherein the R2 group retroelement comprises a R4-AC derived retroelement.

[0365] Embodiment I-8. The nucleic acid of any one of embodiments I-1 to I-7, wherein the engineered protein comprises the at least one heterologous polypeptide fused to the N-terminus of the retroelement-derived polypeptide, to the C-terminus of the retroelement-derived polypeptide and/or internally within the retroelement-derived polypeptide.

[0366] Embodiment I-9. The nucleic acid of embodiment I-8, wherein the engineered protein comprises the at least one heterologous polypeptide fused to the N-terminus of the retroelement-derived polypeptide.

[0367] Embodiment I-10. The nucleic acid of embodiment I-8 or I-9, wherein the engineered protein comprises the at least one heterologous polypeptide fused to the C-terminus of the retroelement-derived polypeptide.

[0368] Embodiment I-11. The nucleic acid of any one of embodiments I-8 to I-10, wherein the engineered protein comprises the at least one heterologous polypeptide fused internally within the retroelement-derived polypeptide.

[0369] Embodiment I-12. The nucleic acid of any one of embodiments I-8 to I-11, wherein the engineered protein comprises a first heterologous polypeptide fused to the N-terminus of the retroelement-derived polypeptide, and a second heterologous polypeptide fused to the C-terminus of the retroelement-derived polypeptide.

[0370] Embodiment I-13. The nucleic acid of any prior embodiment, wherein the at least one heterologous polypeptide comprises an RNA/DNA processing polypeptide.

[0371] Embodiment I-14. The nucleic acid of embodiment I-13, wherein the RNA/DNA processing polypeptide is a RAD51 polypeptide, a RAD17 polypeptide, or a RAD6 polypeptide.

[0372] Embodiment I-15. The nucleic acid of embodiment I-14, wherein the RNA/DNA processing polypeptide is a RAD51 polypeptide having an amino acid sequence that is at least 70% identical to the sequence set forth in SEQ ID NO: 21.

[0373] Embodiment I-16. The nucleic acid of any preceding embodiment, wherein the at least one heterologous polypeptide comprises an RNA/DNA repair polypeptide.

[0374] Embodiment I-17. The nucleic acid of embodiment I-16, wherein the RNA/DNA repair polypeptide is a RAD17 polypeptide, ANKRD28 polypeptide, an HSV-1 alkaline nuclease polypeptide, a BRCA2-derived polypeptide, a PCNA polypeptide, a MDC1 polypeptide, a MSH4 polypeptide, a SCML1 polypeptide, a CDKN2A polypeptide, or a p53 inhibitor.

[0375] Embodiment I-18. The nucleic acid of embodiment I-17, wherein the RNA/DNA repair polypeptide is a RAD17 polypeptide having an amino acid sequence that is at least 70% identical to the sequence set forth in SEQ ID NO: 34.

[0376] Embodiment I-19. The nucleic acid of embodiment I-17, wherein the RNA/DNA repair polypeptide is an ANKRD28 polypeptide having an amino acid sequence that is at least 70% identical to the sequence set forth in SEQ ID NO: 31.

[0377] Embodiment I-20. The nucleic acid of embodiment I-17, wherein the RNA/DNA repair polypeptide is an HSV-1 alkaline nuclease polypeptide having an amino acid sequence that is at least 70% identical to the sequence set forth in SEQ ID NO: 22.

[0378] Embodiment I-21. The nucleic acid of embodiment I-17, wherein the RNA/DNA repair polypeptide is a BRCA2-derived polypeptide having an amino acid sequence that is at least 70% identical to the sequence set forth in SEQ ID NO: 55.

[0379] Embodiment I-22. The nucleic acid of embodiment I-17, wherein the RNA/DNA repair polypeptide is a PCNA polypeptide having an amino acid sequence that is at least 70% identical to the sequence set forth in SEQ ID NO: 36 or SEQ ID NO: 37.

[0380] Embodiment I-23. The nucleic acid of embodiment I-17, wherein the RNA/DNA repair polypeptide is a MDC1 polypeptide having an amino acid sequence that is at least 70% identical to the sequence set forth in SEQ ID NO: 32.

[0381] Embodiment I-24. The nucleic acid of embodiment I-17, wherein the RNA/DNA repair polypeptide is a MSH4 polypeptide having an amino acid sequence that is at least 70% identical to the sequence set forth in SEQ ID NO: 33.

[0382] Embodiment I-25. The nucleic acid of embodiment I-17, wherein the RNA/DNA repair polypeptide is a SCML1 polypeptide having an amino acid sequence that is at least 70% identical to the sequence set forth in SEQ ID NO: 35.

[0383] Embodiment I-26. The nucleic acid of embodiment I-17, wherein the RNA/DNA repair polypeptide is a CDKN2A polypeptide having an amino acid sequence that is at least 70% identical to the sequence set forth in SEQ ID NO: 36.

[0384] Embodiment I-27. The nucleic acid of embodiment I-17, wherein the RNA/DNA repair polypeptide is a p53 inhibitor.

[0385] Embodiment I-28. The nucleic acid of embodiment I-27, wherein the p53 inhibitor is a MDM2-derived peptide having an amino acid sequence that is at least 70% identical to the sequence set forth in SEQ ID NO: 28, or a peptide 14-derived peptide having an amino acid sequence that is at least 70% identical to the sequence set forth in SEQ ID NO: 56.

[0386] Embodiment I-29. The nucleic acid of any preceding embodiment, wherein the at least one heterologous polypeptide comprises a nucleic acid binding polypeptide.

[0387] Embodiment I-30. The nucleic acid of embodiment I-29, wherein the nucleic acid binding polypeptide comprises a non-sequence specific DNA binding polypeptide.

[0388] Embodiment I-31. The nucleic acid of embodiment I-30, wherein the non-sequence specific DNA binding polypeptide comprises a Sto7d DNA binding domain having an amino

acid sequence that is at least 70% identical to the sequence set forth in SEQ ID NO: 42 or an Sso7d DNA binding domain having an amino acid sequence that is at least 70% identical to the sequence set forth in SEQ ID NO: 23.

[0389] Embodiment I-32. The nucleic acid of embodiment I-29, wherein the nucleic acid binding polypeptide comprises a sequence specific DNA binding polypeptide.

[0390] Embodiment I-33. The nucleic acid of embodiment I-32, wherein the sequence specific DNA binding polypeptide is a Cas9 nuclease, dead Cas nuclease, SpCas9 having D10A and/or H840A amino acid substitutions, a Zinc finger DNA binding domain, or a transcription activator-like effector (TALE) DNA binding domain, having an amino acid sequence that is at least 70% identical to the sequence set forth in any one of SEQ ID NOS: 43-54, 60, 80, 81, 82.

[0391] Embodiment I-34. The nucleic acid of any preceding embodiment, wherein at least one heterologous polypeptide comprises a nucleosome binding polypeptide.

[0392] Embodiment I-35. The nucleic acid of embodiment I-34, wherein the nucleosome binding polypeptide comprises an HMGN1 polypeptide having an amino acid sequence that is at least 70% identical to the sequence set forth in SEQ ID NO: 40.

[0393] Embodiment I-36. The nucleic acid of embodiment I-34, wherein the nucleosome binding polypeptide comprises an HMGB1 polypeptide, having an amino acid sequence s that is at least 70% identical to the sequence et forth in SEQ ID NO: 41.

[0394] Embodiment I-37. The nucleic acid of embodiment I-34, wherein the nucleosome binding polypeptide comprises an StkC DNA binding domain, having an amino acid sequence that is at least 70% identical to the sequence set forth in SEQ ID NO: 57.

[0395] Embodiment I-38. The nucleic acid of any preceding embodiment, wherein the engineered polypeptide further comprises a localization signal.

[0396] Embodiment I-39. The nucleic acid of embodiment I-38, wherein the localization signal is a nuclear localization signal (NLS).

[0397] Embodiment I-40. The nucleic acid of embodiment I-39, wherein the NLS comprises an amino acid sequence set forth in any one of SEQ ID NOS: 110-112 or 116.

[0398] Embodiment I-41. The nucleic acid of embodiment I-38, wherein the localization signal is a nucleolar localization signal (NoLS).

[0399] Embodiment I-42. The nucleic acid of embodiment I-41, wherein the NoLS comprises an amino acid sequence set forth in any one of SEQ ID NOs: 113-115.

[0400] Embodiment I-43. The nucleic acid of any preceding embodiment, wherein the engineered protein comprises at least one linker.

[0401] Embodiment I-44. The nucleic acid of embodiment I-43, wherein the linker is at the C-terminus of an N-terminal heterologous polypeptide.

[0402] Embodiment I-45. The nucleic acid of embodiment I-43, wherein the linker is at the N-terminus of a C-terminal heterologous polypeptide.

[0403] Embodiment I-46. The nucleic acid of any one of embodiments I-43 to I-45, wherein the linker is a rigid linker.

[0404] Embodiment I-47. The nucleic acid of any one of embodiments I-43 to I-45, wherein the linker is a flexible linker.

[0405] Embodiment I-48. The nucleic acid of embodiment I-43, wherein the linker is a glycine-serine based linker or a XTEN peptide linker.

[0406] Embodiment I-49. The nucleic acid of any one of embodiments I-43 to I-48, wherein the linker is 2-35 amino acids long and/or wherein the linker is selected from one of SEQ ID NOs: 102-104.

[0407] Embodiment I-50. The nucleic acid of any one of embodiments I-1 to I-49, wherein the nucleic acid encodes an amino acid having an amino acid sequence set forth in any one of SEQ ID NOs: 120-156, 250-267, or 481-483 or having an amino acid sequence that is at least 70% identical to any of SEQ ID NOs: 120-156, 250-260, or 481-483.

[0408] Embodiment I-51. The nucleic acid of embodiment I-50, wherein the nucleic acid encodes an amino acid having an amino acid sequence set forth in SEQ ID NOs: 130-131 or having an amino acid sequence that is at least 70% identical to any of SEQ ID NOs: 130 or 131.

[0409] Embodiment I-52. The nucleic acid of any preceding embodiment, wherein the retroelement-derived polypeptide comprises a reverse transcriptase domain.

[0410] Embodiment I-53. The nucleic acid of embodiment I-52, wherein the reverse transcriptase domain comprises a heterologous reverse transcriptase domain.

[0411] Embodiment I-54. The nucleic acid of embodiment I-53, wherein the heterologous reverse transcriptase domain comprises an amino acid sequence set forth in any one of SEQ ID NOs: 290-301 or having an amino acid sequence that is at least 70% identical to the sequence set forth in in any one of SEQ ID NOs: 290-301.

[0412] Embodiment I-55. The nucleic acid of any preceding embodiment, wherein the retroelement-derived polypeptide comprises a PCNA motif.

[0413] Embodiment I-56. The nucleic acid of embodiment I-55, wherein the PCNA motif comprises a heterologous PCNA motif.

[0414] Embodiment I-57. The nucleic acid of embodiment I-56, wherein the heterologous PCNA motif comprises an amino acid sequence set forth in any one of SEQ ID NOs: 24, 25 or 37 or having an amino acid sequence that is at least 70% identical to the sequence set forth in in any one of SEQ ID NOs: 24, 25 or 37.

[0415] Embodiment I-58. The nucleic acid of any preceding embodiment, wherein the retroelement-derived polypeptide comprises an endonuclease domain.

[0416] Embodiment I-59. The nucleic acid of embodiment I-58, wherein the endonuclease domain comprises a heterologous endonuclease domain.

[0417] Embodiment I-60. The nucleic acid of embodiment I-59, wherein the heterologous endonuclease domain comprises a Cas9 nuclease, a Cas9 nickase, a SpCas9 with H840A mutation, a homing endonuclease, or a FokI nuclease.

[0418] Embodiment I-61. The nucleic acid of embodiment I-60, wherein the heterologous endonuclease domain comprises a Cas9 nuclease having an amino acid sequence set forth in any one of SEQ ID NOs: 80-82 or having an amino acid sequence that is at least 70% identical to the sequence set forth in in any one of SEQ ID NOs: 80-82.

[0419] Embodiment I-62. The nucleic acid of embodiment I-59, wherein the heterologous endonuclease domain comprises SpuFz1 (Fanzor) nuclease.

[0420] Embodiment I-63. The nucleic acid of embodiment I-62, wherein the nucleic acid encodes an amino acid sequence having an amino acid sequence set forth in any one of SEQ ID NOs: 160-171 or having an amino acid sequence that is at least 70% identical to any of SEQ ID NOs: 160-171.

[0421] Embodiment I-64. The nucleic acid of any preceding embodiment, encoding a retroelement polypeptide comprising at least one amino acid substitution relative to an unsubstituted (reference) retroelement polypeptide.

[0422] Embodiment I-65. The nucleic acid of embodiment I-64, wherein the at least one amino acid substitution stabilizes the reverse transcriptase domain and/or its association with RNA and/or DNA relative to an unsubstituted (reference) retroelement reverse transcriptase domain.

[0423] Embodiment I-66. The nucleic acid of embodiment I-64 or I-65, wherein the at least one amino acid substitution is relative to SEQ ID NO: 1.

[0424] Embodiment I-67. The nucleic acid of embodiment I-66, wherein the at least one amino acid substitution is a substitution selected from E399M, Y477H, D555K, W762R, R822K, H1090Y, F956I, P78V, W677C, K314R, K385M, P437M, T448Y, C518Y, T305R, P280K, N269R, L213K, E214K, Q215K, E214R, T1011R, K547R, C145G, I1219N, G257M, K700D, F148K, F667G, F667Y relative to SEQ ID NO: 1.

[0425] Embodiment I-68. The nucleic acid of embodiment I-67, wherein the at least one amino acid substitution is a substitution selected from E399M, Y477H, D555K, W762R, R822K, H1090Y, F956I, P78V, W677C, K314R, K385M, P437M, T448Y, C518Y, T305R, P280K, N269R, L213K, E214K, Q215K, E214R and T1011R relative to SEQ ID NO: 1.

[0426] Embodiment I-69. The nucleic acid of embodiment I-68, wherein the at least one amino acid substitution is a substitution in at least one of positions F148, K700, F667, C145, P280, D555, T1011 relative to SEQ ID NO:1.

[0427] Embodiment I-70. The nucleic acid of embodiment I-68 wherein the at least one amino acid substitution is a substitution in position P280 relative to SEQ ID NO:1.

[0428] Embodiment I-71. The nucleic acid of embodiment I-70, wherein the at least one amino acid substitution is a substitution in position D555 relative to SEQ ID NO:1.

[0429] Embodiment I-72. The nucleic acid of embodiment I-70, wherein the at least one amino acid substitution is a substitution in position T1011 relative to SEQ ID NO:1.

[0430] Embodiment I-73. The nucleic acid of embodiment I-70, wherein the at least one amino acid substitution is a substitution in position K700 relative to SEQ ID NO:1.

[0431] Embodiment I-74. The nucleic acid of embodiment I-66, encoding a retroelement polypeptide having an amino acid sequence set forth in any one of SEQ ID NOs: 230-238.

[0432] Embodiment I-75. The nucleic acid of embodiment I-54 or I-65, wherein the at least one amino acid substitution is relative to SEQ ID NO: 2.

[0433] Embodiment I-76. The nucleic acid of embodiment I-75, wherein the at least one amino acid substitution is a substitution selected from G1057D, S931R, A289P, V996S, L1100F, M287A, C482S, M685L, S547P, V916I, A1256E, M1259A, N332G, M358L, Y919P, T950A, M376P, P151K, D584L, Y1162Q, M640S, C314L, R1131G, A965S, Y1044L, D1236G, A970L, Q1039R, C474Y and T404 relative to SEQ ID NO: 2.

[0434] Embodiment I-77. The nucleic acid of embodiment I-76, wherein the at least one amino acid substitution is a substitution in at least one of positions S931, Q1039 or M287 relative to SEQ ID NO:2.

[0435] Embodiment I-78. The nucleic acid of embodiment I-64 or I-65, wherein the at least one amino acid substitution is relative to SEQ ID NO: 3.

[0436] Embodiment I-79. The nucleic acid of embodiment I-78, wherein the at least one amino acid substitution is a substitution selected from Q195K, W303K, C579P, I629L, A877T, V790D, T866A, R1057N, L1011V, F233Y, A153E, C579S, D124N, Q118K, T410S, H1015V, P1098S, V1029T, T1034R, R1068K, T581N, V790M, V790A, R266I, T410K, D814S, T1038V, I986L, K62M, K326L, or C966G relative to SEQ ID NO: 3.

[0437] Embodiment I-80. The nucleic acid of embodiment I-79, wherein the at least one amino acid substitution is a substitution in at least one of positions Q195 or A877 relative to SEQ ID NO: 3.

[0438] Embodiment I-81. The nucleic acid of any one of embodiments I-1 to I-50, encoding a retroelement polypeptide having an amino acid sequence set forth in any one of SEQ ID NOS: 401-480 or an amino acid sequence that is at least 80% identical or about 85%, 90%, 95%, 98%, or 99% identical to the amino acid sequence set forth in any one of SEQ ID NOS: 401-480.

[0439] Embodiment I-82. The nucleic acid of any preceding embodiment, wherein the nucleic acid is an RNA molecule, a DNA molecule or a hybrid RNA/DNA molecule.

[0440] Embodiment I-83. The nucleic acid of embodiment I-82, wherein the nucleic acid encodes a promoter.

[0441] Embodiment I-84. The nucleic acid of embodiment I-83, wherein the promoter is a heterologous promoter.

[0442] Embodiment I-85. The nucleic acid of embodiment I-83 or I-84, wherein the promoter is a constitutive promoter.

[0443] Embodiment I-86. The nucleic acid of embodiment I-83 or I-84, wherein the promoter is an inducible promoter.

[0444] Embodiment I-87. The nucleic acid of embodiment I-83 or I-84, wherein the promoter is a tissue specific promoter.

[0445] Embodiment I-88. The nucleic acid of embodiment I-83 or I-84, wherein the promoter is selected from an EF1a promoter, a CMV promoter, an A1AT promoter, an Albumin gene promoter, an MNDopt promoter, an MNDu promoter, an SFFV promoter, or an ApoE promoter.

[0446] Embodiment I-89. The nucleic acid of any preceding embodiment, wherein the nucleic acid comprises one or more chemical or sequence modifications.

[0447] Embodiment I-90. The nucleic acid of embodiment I-89, wherein the one or more chemical or sequence modifications are selected from an RNA CAP, a modified polyA length, a

chemical modification, a 5' UTR modification, a 3' UTR modification, a modified Kozak sequence, a modified (e.g., truncated) stem loop, an RNA stabilization motif, a 5-methoxyuridine (5-MO-U) modification, a 5-methylcytidine (5mC) modification, or one or more additional and/or modified microsatellites.

[0448] Embodiment I-91. The nucleic acid of embodiment I-90, wherein the RNA stabilization motif is a WPRE motif.

[0449] Embodiment I-92. The nucleic acid of any preceding embodiment comprising a codon optimized sequence for expression in human cells.

[0450] Embodiment I-93. The nucleic acid of embodiment I-92, wherein the codon optimized sequence comprises one or more of a reduced Uracil (U) load relative to a corresponding naturally occurring sequence, reduced GC% content relative to a corresponding naturally occurring sequence, reduced the length and/or number of intron sequences relative to a corresponding naturally occurring sequence, reduced number of RNA binding motifs or sites relative to a corresponding naturally occurring sequence, lower ΔG (free energy) relative to a corresponding naturally occurring sequence, reduced number of nucleotide repeats relative to a corresponding naturally occurring sequence, reduced number of palindromic sequences relative to a corresponding naturally occurring sequence, maximized pairing of bases relative to a corresponding naturally occurring sequence, removal of splicing site sequences and/or removal of rare codons or other slowly translated codons.

[0451] Embodiment I-94. A nucleic acid comprising a transgene and an expression cassette for a cis-acting RNA binding protein configured to be translated in cis with the transgene.

[0452] Embodiment I-95. The nucleic acid of embodiment I-94, wherein the RNA-binding activity is derived from the ORF1p of a retrotransposable element.

[0453] Embodiment I-96. The nucleic acid of embodiment I-95, where the ORF1p derives from ZFL2-1 retrotransposable element from Zebrafish, from DOC2 retrotransposable element from Drosophila, from Maui retrotransposable element from Fugu fish I-3 retrotransposable element from Zebrafish.

[0454] Embodiment I-97. The nucleic acid of embodiment I-95 or I-96, comprising a nucleic acid sequence set forth in any one of SEQ ID NOs: 190-204 or at least 70% identical to any of SEQ ID NOs: 190-204.

[0455] Embodiment I-98. A nucleic acid encoding a retrotransposable element comprising an amino acid sequence set forth in any one of SEQ ID NOs: 401-480.

[0456] Embodiment I-99. An engineered protein encoded by any one of the nucleic acids of any one of embodiments I-1 to I-98.

[0457] Embodiment I-100. A composition comprising:

- a) a first nucleic acid of any one of embodiments I-1 to I-93; and
- b) a second nucleic acid comprising a polynucleotide encoding a gene of interest.

[0458] Embodiment I-101. The composition of embodiment I-100, wherein the first and second nucleic acids are separate DNA polynucleotides.

[0459] Embodiment I-102. The composition of embodiment I-100, wherein the first and second nucleic acids are separate RNA polynucleotides.

[0460] Embodiment I-103. The composition of embodiment I-100, wherein one of the first and second nucleic acids is a DNA polynucleotide and one of the first and second nucleic acids is an RNA polynucleotide.

[0461] Embodiment I-104. The composition of any one of embodiments I-100 to I-103, wherein the first nucleic acid is operably linked to a first heterologous promoter and/or wherein the second nucleic acid is operably linked to a second heterologous promoter.

[0462] Embodiment I-105. The composition of any one of embodiments I-100 to I-104, wherein one or both of the first and second nucleic acids further comprise one or more modification selected from an RNA CAP, a modified polyA length, a chemical modification (e.g., a pseudouridine and/or a methylpseudouridine), a 5' UTR modification, a 3' UTR modification, a modified Kozak sequence, a modified stem loop, an RNA stabilization motif, a 5-

methoxyuridine (5-MO-U) modification, a 5-methylcytidine (5mC) modification, or one or more additional and/or modified microsatellites.

[0463] Embodiment I-106. The composition of any one of embodiments I-100 to I-104, wherein one or both of the first and second nucleic acids comprises a codon optimized sequence.

[0464] Embodiment I-107. The composition of embodiment I-106, wherein the codon optimized sequence is optimized for expression in human cells.

[0465] Embodiment I-108. The composition of any one of embodiments I-100 to I-107, wherein one or both of the first and second nucleic acids further comprise an RNA nuclear localization sequence.

[0466] Embodiment I-109. The composition of any one of embodiments I-100 to I-108, wherein the polynucleotide encoding the gene of interest in the second nucleic acid is flanked by a first terminal region and a second terminal region.

[0467] Embodiment I-110. The composition of embodiment I-109, wherein the second nucleic acid comprises flanking regions homologous to target sites in a genome of a target cell.

[0468] Embodiment I-111. The composition of any one of embodiments I-100 to I-110, wherein the second nucleic acid comprises i) a promoter operably linked to the polynucleotide encoding the gene of interest, and ii) a polyadenylation sequence, and wherein the promoter is selectively active in one or more target cell types.

[0469] Embodiment I-112. The composition of embodiment I-110, wherein the target site is a genome of a target cell.

[0470] Embodiment I-113. The composition of any one of embodiments I-100 to I-112, wherein the gene of interest encodes a therapeutic RNA and/or a therapeutic protein.

[0471] Embodiment I-114. The composition of embodiment I-113, wherein the therapeutic RNA is an antisense RNA (asRNA), small interfering RNA (siRNA), microRNA (miRNA), or RNA aptamer.

[0472] Embodiment I-115. The composition of embodiment I-113, wherein the therapeutic protein is an antibody, regulatory protein, hormone, cytokine, structural protein, enzyme, or membrane protein.

[0473] Embodiment I-116. The composition of embodiment I-113, wherein the therapeutic protein is Factor VIII, Factor IX, Phenylalanine hydroxylase, ATP7B, alpha glucosidase, arginosuccinate synthetase, galactose-1-phosphate uridylyltransferase, or ornithine transcarbamylase.

[0474] Embodiment I-117. The composition of any one of embodiments I-100 to I-116, wherein the first and second nucleic acids are comprised within a plurality of LNP particles.

[0475] Embodiment I-118. A composition comprising the nucleic acids of any one of embodiments I-94 to I-97.

[0476] Embodiment I-119. The composition of embodiment I-118, wherein the nucleic acid is comprised within a plurality of LNP particles.

[0477] Embodiment I-120. A method of treating a subject in need thereof, comprising administering to the subject a nucleic acid of any one of embodiments I-1 to I-98, an engineered protein of embodiment I-99 or a composition of any one of embodiments I-100 to I-119, to a subject.

[0478] Embodiment I-121. Use of a nucleic acid of any one of embodiments I-1 to I-98, an engineered protein of embodiment I-99 or a composition of any one of embodiments I-100 to I-119 to treat a subject in need thereof.

[0479] Embodiment I-122. The method of embodiment I-120 or use of embodiment I-121, wherein the subject is a mammal.

[0480] Embodiment I-123. The method or use of embodiment I-122, wherein the mammal is a human.

[0481] Embodiment I-124. The composition of embodiment I-109, wherein the first terminal region comprises a sequence set forth in any one of SEQ ID NOS: 215, 217, 219, and 500-605.

[0482] Embodiment I-125. The composition of embodiment I-109, wherein the first terminal region comprises a sequence set forth in any one of SEQ ID NOS: 216, 218, 220, and 606-712.

Set II:

[0483] Some embodiments of this disclosure relate to Embodiment set II, as follows:

[0484] Embodiment II-1. A nucleic acid encoding an engineered driver comprising an R2 protein or portion thereof fused to at least one heterologous polypeptide, wherein:

a. the at least one heterologous polypeptide comprises a heterologous nucleosome binding polypeptide, a heterologous nucleic acid binding polypeptide, a heterologous RNA/DNA repair polypeptide, a heterologous RNA/DNA processing polypeptide, a heterologous localization signal, a heterologous reverse transcriptase, a heterologous endonuclease, a heterologous intrinsically disordered polypeptide, or any combination thereof; and

b. the engineered driver exhibits an improvement in at least efficiency, compared to a reference R2 protein.

[0485] Embodiment II-2. The nucleic acid of embodiment II-1, wherein the at least one heterologous polypeptide is fused to the N-terminus of the R2 protein or portion thereof.

[0486] Embodiment II-3. The nucleic acid of embodiment II-1, wherein the at least one heterologous polypeptide is fused to the C-terminus of the R2 protein or portion.

[0487] Embodiment II-4. The nucleic acid of embodiment II-1, wherein the at least one heterologous polypeptide is fused internally within the R2 protein or portion thereof, or a combination thereof.

[0488] Embodiment II-5. The nucleic acid of any one of embodiments II-1 to II-4, wherein at the least one heterologous polypeptide is a heterologous nucleosome binding polypeptide.

[0489] Embodiment II-6. The nucleic acid of embodiment II-5, wherein the heterologous nucleosome binding polypeptide is a HMGN1 polypeptide, a HMGB1 polypeptide, or a StkC DNA binding domain.

[0490] Embodiment II-7. The nucleic acid of embodiment II-6, wherein the heterologous nucleosome binding polypeptide is a HMGN1 polypeptide.

[0491] Embodiment II-8. The nucleic acid of embodiment II-7, wherein the HMGN1 polypeptide comprises an amino acid sequence set forth in SEQ ID NO: 40, or an amino acid sequence that is at least 70% identical thereto.

[0492] Embodiment II-9. The nucleic acid of embodiment II-8, wherein the HMGN1 polypeptide comprises at least one amino acid substitution, wherein the at least one amino acid substitution is: D100E or D100G, relative to SEQ ID NO: 40.

[0493] Embodiment II-10. The nucleic acid of any one of embodiments II-1 to II-9, wherein the at least one heterologous polypeptide comprises a heterologous RNA/DNA repair polypeptide.

[0494] Embodiment II-11. The nucleic acid of embodiment II-10, wherein the at least one heterologous RNA/DNA repair polypeptide is a CTIP polypeptide.

[0495] Embodiment II-12. The nucleic acid of embodiment II-10, wherein the at least one heterologous RNA/DNA repair polypeptide comprises a CTIP polypeptide.

[0496] Embodiment II-13. The nucleic acid of embodiment II-11 or II-12, wherein the CTIP polypeptide comprises an amino acid sequence set forth in SEQ ID NO: 61 or an amino acid sequence that is at least 70% identical thereto.

[0497] Embodiment II-14. The nucleic acid of embodiment II-13, wherein the CTIP polypeptide comprises at least one amino acid substitution, wherein the at least one amino acid substitution is: T15I, R123K, P181L, M202L, M202F, M202T, A237L, A237I, A237N, A237K, N288D, D18F, I49C, S17D, D18C, Q46T, I2A, C198S, E155S, L281G, I186S, Q46K, or a combination thereof relative to SEQ ID NO: 61.

[0498] Embodiment II-15. The nucleic acid of embodiment II-10, wherein the at least one heterologous RNA/DNA repair polypeptide comprises a heterologous PIP motif polypeptide, and the heterologous PIP motif polypeptide is p21, FEN1 or CHAF1A.

[0499] Embodiment II-16. The nucleic acid of embodiment II-15, wherein, the R2 protein or portion thereof lacks a functional native PIP motif.

[0500] Embodiment II-17. The nucleic acid of any one of embodiments II-1 to II-16, wherein the at least one heterologous polypeptide comprises a heterologous RNA/DNA processing polypeptide.

[0501] Embodiment II-18. The nucleic acid of embodiment II-17, wherein the heterologous RNA/DNA processing polypeptide is a RNA helicase polypeptide, a RNA helicase recruitment motif polypeptide, or a RAD51 polypeptide.

[0502] Embodiment II-19. The nucleic acid of embodiment II-18, wherein the heterologous RNA/DNA processing polypeptide is a RNA helicase polypeptide.

[0503] Embodiment II-20. The nucleic acid of embodiment II-19, wherein the RNA helicase polypeptide is a EIF4A polypeptide.

[0504] Embodiment II-21. The nucleic acid of embodiment II-20, wherein the EIF4A polypeptide comprises an amino acid sequence set forth in SEQ ID NO: 585, or an amino acid sequence that is at least 70% identical thereto.

[0505] Embodiment II-22. The nucleic acid of embodiment II-19, wherein the RNA helicase polypeptide is a Dengue virus RNA helicase polypeptide.

[0506] Embodiment II-23. The nucleic acid of embodiment II-22, wherein the Dengue virus RNA helicase polypeptide comprises an amino acid sequence set forth in SEQ ID NO: 1875, or an amino acid sequence that is at least 70% identical thereto.

[0507] Embodiment II-24. The nucleic acid of embodiment II-18, wherein the at least one heterologous polypeptide comprises a RNA helicase recruitment motif polypeptide.

[0508] Embodiment II-25. The nucleic acid of embodiment II-24, wherein the RNA helicase recruitment motif polypeptide is a TUDOR domain polypeptide.

[0509] Embodiment II-26. The nucleic acid of embodiment II-25, wherein the TUDOR domain polypeptide comprises an amino acid sequence set forth in SEQ ID NO: 584, or an amino acid sequence that is at least 70% identical thereto.

[0510] Embodiment II-27. The nucleic acid of embodiment II-26, wherein the TUDOR domain polypeptide comprises an amino acid sequence set forth in SEQ ID NO: 1853, or an amino acid sequence that is at least 70% identical thereto.

[0511] Embodiment II-28. The nucleic acid of embodiment II-24, wherein the RNA helicase recruitment motif polypeptide is a canine hepacivirus core protein polypeptide.

[0512] Embodiment II-29. The nucleic acid of embodiment II-28, wherein the canine hepacivirus core protein polypeptide comprises an amino acid sequence set forth in SEQ ID NO: 789, or an amino acid sequence that is at least 70% identical thereto.

[0513] Embodiment II-30. The nucleic acid of embodiment II-18, wherein the at least one heterologous polypeptide comprises a RAD51 polypeptide.

[0514] Embodiment II-31. The nucleic acid of embodiment II-30, wherein the RAD51 polypeptide comprises an amino acid sequence set forth in SEQ ID NO: 21, or an amino acid sequence that is at least 70% identical thereto.

[0515] Embodiment II-32. The nucleic acid of any one of embodiments II-1 to II-31, wherein the at least one heterologous polypeptide comprises a heterologous nucleic acid binding polypeptide.

[0516] Embodiment II-33. The nucleic acid of embodiment II-32, wherein the heterologous nucleic acid binding polypeptide is a non-sequence specific DNA binding polypeptide.

[0517] Embodiment II-34. The nucleic acid of embodiment II-33, wherein the non-sequence specific DNA binding polypeptide comprises a Sto7d DNA binding domain polypeptide, a Sso7d DNA binding domain polypeptide, or a UL42 DNA binding domain polypeptide.

[0518] Embodiment II-35. The nucleic acid of embodiment II-34, wherein the Sto7d DNA binding domain polypeptide comprises an amino acid sequence set forth in SEQ ID NO: 42, or an amino acid sequence that is at least 70% identical thereto.

[0519] Embodiment II-36. The nucleic acid of embodiment II-34, wherein the Sso7d DNA binding domain polypeptide comprises an amino acid sequence set forth in SEQ ID NO: 23, or an amino acid sequence that is at least 70% identical thereto.

[0520] Embodiment II-37. The nucleic acid of embodiment II-34, wherein the UL42 DNA binding domain polypeptide comprises an amino acid sequence set forth in SEQ ID NO: 587 or an amino acid sequence that is at least 70% identical thereto.

[0521] Embodiment II-38. The nucleic acid of any one of embodiments II-1 to II-37, wherein the at least one heterologous polypeptide is a heterologous intrinsically disordered region polypeptide.

[0522] Embodiment II-39. The nucleic acid of embodiment II-38, wherein the heterologous intrinsically disordered region polypeptide comprises an amino acid sequence set forth in any one of SEQ ID NO: 1171-1216 or an amino acid sequence that is at least 70% identical thereto.

[0523] Embodiment II-40. The nucleic acid of embodiment II-38 or II-39, wherein the R2 protein or portion thereof has a deletion in a native intrinsically disordered region.

[0524] Embodiment II-41. The nucleic acid of any one of embodiments II-1 to II-40, wherein the at least one heterologous polypeptide comprises a heterologous localization signal.

[0525] Embodiment II-42. The nucleic acid of embodiment II-41, wherein the heterologous localization signal is a nuclear localization signal (NLS).

[0526] Embodiment II-43. The nucleic acid of embodiment II-42, wherein the NLS comprises an amino acid sequence set forth in any one of SEQ ID NOs: 110-112, 116, or 1854.

[0527] Embodiment II-44. The nucleic acid of embodiment II-41, wherein the localization signal is a nucleolar localization signal (NoLS).

[0528] Embodiment II-45. The nucleic acid of embodiment II-44, wherein the NoLS comprises an amino acid sequence set forth in any one of SEQ ID NOs: 113-115.

[0529] Embodiment II-46. The nucleic acid of any one of embodiments II-1 to II-45, wherein the R2 protein or portion thereof comprises a reverse transcriptase domain and an endonuclease domain.

[0530] Embodiment II-47. The nucleic acid of any one of embodiments II-1 to II-46, wherein the R2 protein is a wild type (WT) R2 protein.

[0531] Embodiment II-48. The nucleic acid of any one of embodiments II-1 to II-46, wherein the R2 protein or portion thereof comprises at least one amino acid substitution, at least one amino acid deletion, at least one amino acid addition, or a combination thereof, relative to a WT R2 protein.

[0532] Embodiment II-49. The nucleic acid of embodiment II-48, wherein the R2 protein or portion thereof has at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% sequence identity with the WT R2 protein.

[0533] Embodiment II-50. The nucleic acid of any one of embodiments II-47 to II-49, wherein the WT of the R2 protein comprises an amino acid sequence set forth in any one of SEQ ID NOS: 1-2, 401-480, and 30001-30060.

[0534] Embodiment II-51. The nucleic acid of any one of embodiments II-48 to II-50, wherein the R2 protein or portion thereof comprises at least one amino acid substitution, relative to the WT of the R2 protein.

[0535] Embodiment II-52. The nucleic acid of embodiment II-51, wherein the at least one amino acid substitution stabilizes the reverse transcriptase domain of the R2 protein and/or the association of the reverse transcriptase domain with an RNA and/or a DNA, relative to the reverse transcriptase domain of the WT of the R2 protein.

[0536] Embodiment II-53. The nucleic acid of embodiment II-51 or II-52, wherein the at least one amino acid substitution is relative to SEQ ID NO: 1.

[0537] Embodiment II-54. The nucleic acid of embodiment II-53, wherein the at least one amino acid substitution is a substitution in at least one of positions F148, K700, F667, C145, P280, D555, T1011, I1219, and A1282 relative to SEQ ID NO:1.

[0538] Embodiment II-55. The nucleic acid of embodiment II-53 or II-54, wherein the at least one amino acid substitution position P280 relative to SEQ ID NO:1.

[0539] Embodiment II-56. The nucleic acid of embodiment II-53 or II-54, wherein the at least one amino acid substitution is a substitution in position D555 relative to SEQ ID NO:1.

[0540] Embodiment II-57. The nucleic acid of embodiment II-53 or II-54, wherein the at least one amino acid substitution is a substitution in position T1011 relative to SEQ ID NO:1.

[0541] Embodiment II-58. The nucleic acid of embodiment II-53 or II-54, wherein the at least one amino acid substitution is a substitution in position I1219 relative to SEQ ID NO:1.

[0542] Embodiment II-59. The nucleic acid of embodiment II-53 or II-54, wherein the at least one amino acid substitution is a substitution in position A1282 relative to SEQ ID NO:1.

[0543] Embodiment II-60. The nucleic acid of embodiment II-53, wherein the at least one amino acid substitution is a substitution selected from the group consisting of L18P, E19L, S26N, V27R, A29T, I30A, R32G, N35I, S36R, L37K, A38R, L49F, G209R, K289R, L316H, E399M, E472G, Y477H, T514R, D537Y, T538K, H544Q, D555K, N562D, W762R, R822K, S895K, Q909N, R964K, G966N, T989Y, I1003M, T1011S, A1012G, L1013R, T1016L, Q1019R, E1027D, P1033R, C1079K, K1085R, L1089R, H1090Y, I1219N, A1282G, T989I, K579R, K547R, V660I, T608P, F956I, I1003L, T1016S, T680S, L594P, P52R, P418S, P78G, L51G, P78V, P180S, V539G, P180K, P437R, V539R, N640A, W677C, V687W, R866I, A747G, W762E, T908I, W959K, D1037N, Y1194L, R1292P, K1210R, K993R, K755R, C518R, K314R, K1182R, K385M, P437M, T448Y, C518Y, E1008R, T305R, P280K, A1012R, N269R, L213K, E214K, Q215K, E214R, T1011R, E1015R, L213R, G257M, G209F, W762D, Y1194L, V687W, W677C, R1292P, T908I, W959K, N640A, Q909N, R866I, A1282G, W762E, W959K, T908I, Y579R, T1011A, Y571H, E630D, D1075E, L1089M, A442P, H423Q, E173D, P188L, C145G, N301G, R355K, G257R, V251G, T603A, N363H, P1033S, F667Y, F1365L, V501L, G257R+C145G, W762R+C145G, G257R+E173D, N363H+E173D, A442P+E173D, W762R+E173D, G257R+P188L, W762R+P188L, G257R+V251G, S895K+V251G, N363H+G257R, R568S, E542K, P985R, G561R, Y563R, L1184K, H1030R, A557R, I558R, Q567R, N271W, E350R, L316R, F304K, D923K, A317W, E1010R, Q1021R, E419R, T1011F, E769R, W345R, V110R, R568G, F135G, N104Y, V142Y, P115K, A111Y, L129R, V98R, D126R, E144R, Q112R, Y136R, F148K, T1016K, K700D, E188D, E188A, R435K, L438A, A1191L, P1195G, P1195G, R1197K, L1200I, G1331A, G1335A, P593G, S595V, E596S, S599V, P593A, S595A, E596A, S599A, K612D, D616K, K612A, D616A, K856D, D6860K, K856A, D6860A, N891S, T892V, S895V, N891A, T892A, S895A, E888S, E898S, T899V, N902S, E888A, E898A, T899A, N902A, T934V, G937P, P939G, T934A, G937A, K1252D, D1255K, W1256G, K1252A, D1255A, W1256A, A877V, A877I, A877M, F876Y, A877H,

Q955R, L952I, T989L, E947S, G658R, Q642R, F973R, G647T, D990V, D990R, N941K, V727L, M874L, D227W, D227F, C344V, R669V, V600T, R866L, T989V, M449E, T679L, Y478L, M671S, F956V, M686S, W1126P, T1367R, C404S, E1363K, F973Q, M420T, R669F, G689E, N1219L, K480R, K508Q, E812K, V1368R, R1157M, G217P, H1320E, N891K, K1269G, K1269G, D1255G, T670L, T670L, H1180Q, K907R, T996A, I778F, R574G, K1151H, E194G, R574S, D1288A, F186G, T969E, G432Q, D591Q, M553R, T174K, W1028K, D534Y, D591W, D972R, W1028Y, D972Q, P169R, L975R, C587K, I978K, D591R, D182R, D972M, A1024K, C170K, P985K, G172R, S178R, G172K, S177R, A166R, S177W, P548R, S204R, H544R, G575H, F1135T, N1123K, P1116D, P998H, S645Q, P998Q, D879A, D879M, K576M, D879W, E643R, T994W, N1140R, D1133R, D583R, M644R, G575R, A1128R, A649R, E1125R, F1135Y, C750K, P1129R, F1135R, L1057R, H1006K, L582K, D573R, T994R, D878R, A578K, D839R, G311S, T532K, T728K, P1116K, L243K, E247K, L249F, K255R, V258I, F260Y, Y272H, E286K, P291A, D303S, I308R, I327L, K333R, G339R, I370L, I374L, K377R, S383K, L474I, T481I, L482I, F498L, M502L, Q516R, F519I, G520K, C521L, S526K, D534K, L572K, S589R, F611Y, P545K, F549R, L564R, L569R, F570Y, D573N, K377R, L582I, I585K, I597V, F601Y, F628L, L631P, V641L, A649S, P650A, V661K, E666K, E672R, F674Y, W677I, V687L, C690S, N738H, Q741R, L754I, A764S, H768R, L771I, V776I, F846Y, L848I, I896L, W933F, Y995F, A1038G, A1063L, I778F, V785I, L794M, E798G, K821G, K821R, V835I, L844I, L854I, C855R, H865K, C901L, D929N, A932P, M942L, K950R, Y951H, D957N, S965P, T969E, P998S, Y1002H, L1014I, T1023K, E1027K, Y1041H, S1043R, K1053R, V1061R, Q1062R, T1130K, Q1173R, Q1233K, F1278L, V1287I, V1291I, E1317Q, V1318I, M1361L, E1363K, R1064K, K1078R, F1080L, E1082K, K1092R, G1098R, E1127K, E1127Q, K1132R, Q1155H, F1162Y, W1171F, L1184I, Q1188K, C1215Y, N1228Q, E1249K, V1257T, H1313P, F1345N, K1355R, K1386R, V727M, T759G, Q795R, I814C, T819E, H825K, Q868G, Q868D, A931P, D944G, F973S, L175S, A1012S, I683L, S699G, I707P, W710Y, T714A, I718T, I726V, N813G, V881T, L906M, S648K, S699K, S717P, A748R, D800P, V239M, E334A, T532G, K622Q, T1367R, H1378R, F1380Y, Y817R, Q911K, Q911S, L657Y, L720Y, A779S, K780N, D199K, A779K, D783G, D851N, E949Y, E949Y, V299E, E949Y, G520A, D591G, S599D, T1122L, L1285I, N1103D, P739R, E638L, N889G, Y1242K, E400M, D1133P, S645K, V639T, I745R, G772A, A747S, A748P, I778F, M841L, M850L, E949L, F983E, R669L, L213K, L564Q, F973E, R669I, F667G, and F973E, or combinations thereof relative to SEQ ID NO: 1.

[0544] Embodiment II-61. The nucleic acid of embodiment II-53, wherein the at least one amino acid substitution is a substitution selected from the group consisting of E399M, Y477H, D555K, W762R, R822K, H1090Y, F956I, P78V, W677C, K314R, K385M, P437M, T448Y, C518Y, T305R, P280K, N269R, L213K, E214K, Q215K, T1011R, R435K, L438A, R1197K, G1331A, T989L, Q955R, L952I, E947S, Q642R, G647T, N941K, V727L, M874L, D227W, D227F, C344V, R669V, V600T, T989V, M449E, T679L, Y478L, M671S, F956V, M686S, W1126P, T1367R, C404S, E1363K, K508Q, E812K, V1368R, R1157M, N1123K, P1116D, P998H, S645Q, P998Q, N1123K, P1116D, P998H, S645Q, P998Q, D879A, D879M, K576M, G311S, T532K, L572K, S589R, F611Y, F846Y, L848I, I896L, W933F, Y995F, A1038G, A1063L, T1130K, Q1173R, Q1233K, F1278L, V1287I, V1291I, E1317Q, V1318I, M1361L, E1363K, K1386R, V727M, T759G, Q795R, I814C, T819E, H825K, Q868G, Q868D, A931P, D944G, F973S, L175S, A1012S, I683L, S699G, I707P, W710Y, T714A, I718T, I726V, N813G, V881T, L906M, S648K, S699K, S717P, A748R, D800P, V239M, E334A, T532G, K622Q, D923K, P739R, E638L, N889G, Y1242K, E400M, T1011S, T1011A, I1219N, A1282G, L213R, A1012G, E542K, G257R, C145G, L316R, D537Y, D1133P, and K1085R, or combinations thereof relative to SEQ ID NO: 1.

[0545] Embodiment II-62. The nucleic acid of embodiment II-53, wherein the at least one amino acid substitution comprises the following substitutions: P280K, D555K, T1011S, I1219N, and A1282G relative to SEQ ID NO:1.

[0546] Embodiment II-63. The nucleic acid of embodiment II-53, wherein the at least one amino acid substitution is a substitution in at least one of positions G257R, P280K, N363H, E399M, V539G, H423Q, A442P, D537Y, K547R, D555K, R669L, N889G, F973E, T1011A, Q1019R, D1133P, K1085R, R1177G, Y1242K, A1282G relative to SEQ ID NO:1.

[0547] Embodiment II-64. The nucleic acid of embodiment II-63, wherein the at least one amino acid substitution comprises the following substitutions: G257R, P280K, N363H, E399M, V539G, H423Q, A442P, D537Y, K547R, D555K, R669L, N889G, F973E, T1011A, Q1019R, D1133P, K1085R, R1177G, Y1242K, and A1282G relative to SEQ ID NO:1

[0548] Embodiment II-65. The nucleic acid of embodiment II-53, wherein the engineered driver comprises an amino acid sequence set forth in any one of SEQ ID NOs: 230-238, 668, and 1378, or an amino acid sequence that is at least 70% identical thereto.

[0549] Embodiment II-66. The nucleic acid of embodiment II-51 or II-52, wherein the at least one amino acid substitution is relative to SEQ ID NO: 2.

[0550] Embodiment II-67. The nucleic acid of embodiment II-66, wherein the at least one amino acid substitution is a substitution selected from the group consisting of G1057D, S931R, A289P, V996S, L1100F, M287A, C482S, M685L, S547P, V916I, A1256E, M1259A, N332G, M358L, Y919P, T950A, M376P, P151K, D584L, Y1162Q, M640S, C314L, R1131G, A965S, Y1044L, D1236G, A970L, Q1039R, C474Y, and T404, or combinations thereof relative to SEQ ID NO: 2.

[0551] Embodiment II-68. The nucleic acid of embodiment II-66, wherein the at least one amino acid substitution is a substitution in at least one of positions S931, Q1039 or M287 relative to SEQ ID NO: 2.

[0552] Embodiment II-69. The nucleic acid of any one of embodiments II-1 to II-46 and II-48 to II-68, wherein the R2 protein or portion thereof comprises at least one amino acid deletion, wherein the at least one amino acid deletion comprises a deletion of a N-terminal intrinsically disordered region or a portion thereof of the R2 protein.

[0553] Embodiment II-70. The nucleic acid of embodiment II-69, wherein the at least one amino acid deletion is a deletion of between about 10 amino acids and about 300 amino acids within the N-terminal intrinsically disordered region.

[0554] Embodiment II-71. The nucleic acid of embodiment II-69, wherein the N-terminal intrinsically disordered region is located at about 10 amino acids and about 300 amino acids within the N-terminus of the WT R2 protein.

[0555] Embodiment II-72. The nucleic acid of any one of embodiments II-69 to II-71, wherein the at least one amino acid deletion is a deletion of between about 10 amino acids and about 300 amino acids within the N-terminus of the R2 protein relative to SEQ ID NO: 1.

[0556] Embodiment II-73. The nucleic acid of any one of embodiments II-69 to II-71, wherein the at least one amino acid deletion is a deletion of between about 160 amino acids and about 240 amino acids at the N-terminus of the R2 protein relative to SEQ ID NO: 1.

[0557] Embodiment II-74. The nucleic acid of any one of embodiments II-69 to II-71, wherein the at least one amino acid deletion is a deletion of about 161 amino acids, about 184 amino acids, or about 218 amino acids, at the N-terminus the R2 protein relative to SEQ ID NO: 1.

[0558] Embodiment II-75. The nucleic acid of any one of embodiments II-1 to II-45 and II-48 to II-74, wherein the R2 protein or portion thereof comprises a native reverse transcriptase domain and does not comprise a functional native endonuclease domain.

[0559] Embodiment II-76. The nucleic acid of embodiment II-75, wherein the at least one heterologous polypeptide comprises a heterologous endonuclease domain.

[0560] Embodiment II-77. The nucleic acid of embodiment II-76, wherein the heterologous endonuclease domain is an endonuclease domain of another R2 protein.

[0561] Embodiment II-78. The nucleic acid of embodiment II-77, wherein the another R2 protein has an amino acid sequence set forth in any one of SEQ ID NOS: 1, 2, 400-480 and 30001-30060, or an amino acid sequence that is at least 70% identical thereto.

[0562] Embodiment II-79. The nucleic acid of any one of embodiments II-1 to II-45 and II-48 to II-74, wherein the R2 protein comprises a native endonuclease domain and does not comprise a functional native reverse transcriptase domain.

[0563] Embodiment II-80. The nucleic acid of embodiment II-79, wherein the at least one heterologous polypeptide comprises the heterologous reverse transcriptase domain.

[0564] Embodiment II-81. The nucleic acid of embodiment II-80, wherein the heterologous reverse transcriptase domain is a reverse transcriptase domain of another WT R2 protein.

[0565] Embodiment II-82. The nucleic acid of embodiment II-81, wherein the another R2 protein has an amino acid sequence set forth in any one of SEQ ID NOS: 1, 2, 400-480 and 30001-30060, or an amino acid sequence that is at least 70% identical thereto.

[0566] Embodiment II-83. The nucleic acid of embodiment II-80, wherein heterologous reverse transcriptase domain comprises an amino acid sequence set forth in SEQ ID NO:295, or an amino acid sequence that is at least 70% identical thereto.

[0567] Embodiment II-84. The nucleic acid of any one of embodiments II-1 to II-83, wherein the nucleic acid is a RNA molecule, a DNA molecule or a hybrid RNA/DNA molecule.

[0568] Embodiment II-85. The nucleic acid of any one of embodiments II-1 to II-84, wherein the nucleic acid comprises one or more chemical or sequence modifications.

[0569] Embodiment II-86. The nucleic acid of embodiment II-85, wherein the one or more chemical or sequence modifications are selected from the group consisting of an RNA CAP, a modified polyA length, a linker, a chemical modification, a 5' UTR modification, a 3' UTR modification, a modified Kozak sequence, a modified (e.g., truncated) stem loop, an RNA stabilization motif, a 5-methoxyuridine (5-MO-U) modification, a 5-methylcytidine (5mC) modification, and one or more additional and/or modified microsatellites.

[0570] Embodiment II-87. The nucleic acid of any one of embodiments II-1 to II-86, wherein the engineered driver comprises a heterologous CTIP polypeptide and a heterologous HMGN1 polypeptide.

[0571] Embodiment II-88. The nucleic acid of embodiment II-87, wherein the heterologous CTIP polypeptide and the heterologous HMGN1 polypeptide are situated on the N-terminal side of the R2 protein or portion thereof.

[0572] Embodiment II-89. The nucleic acid of embodiment II-88, wherein the heterologous CTIP polypeptide is situated on the N-terminal side of the heterologous HMGN1 polypeptide.

[0573] Embodiment II-90. The nucleic acid of any one of embodiments II-87 to II-89, wherein the R2 protein or portion thereof comprises a deletion within a native N-terminal intrinsically disordered region.

[0574] Embodiment II-91. The nucleic acid of any one of embodiments II-87 to II-90, wherein the engineered driver comprises a heterologous intrinsically disordered region.

[0575] Embodiment II-92. The nucleic acid of embodiment II-91, wherein the heterologous intrinsically disordered region is situated on the N-terminal side of the heterologous CTIP polypeptide and the heterologous HMGN1 polypeptide.

[0576] Embodiment II-93. The nucleic acid of embodiment II-92, wherein the engineered driver has an amino acid sequence set forth in any one of SEQ ID NOS: 679-714, or an amino acid sequence that is at least 70% identical thereto.

[0577] Embodiment II-94. The nucleic acid of any one of embodiments II-87 to II-90, wherein the engineered driver comprises at least one heterologous RNA/DNA processing polypeptide.

[0578] Embodiment II-95. The nucleic acid of embodiment II-94, wherein the at least one heterologous RNA/DNA processing polypeptide comprises a RNA helicase polypeptide.

[0579] Embodiment II-96. The nucleic acid of embodiment II-95, wherein the RNA helicase polypeptide is situated at the N-terminal side of the heterologous CTIP polypeptide and the heterologous HMGN1 polypeptide.

[0580] Embodiment II-97. The nucleic acid of embodiment II-96, wherein the engineered driver has an amino acid sequence set forth in SEQ ID NO: 783, or an amino acid sequence that is at least 70% identical thereto.

[0581] Embodiment II-98. The nucleic acid of any one of embodiments II-1 to II-86, wherein the engineered driver comprises a heterologous Sto7d polypeptide and a heterologous HMGN1 polypeptide.

[0582] Embodiment II-99. The nucleic acid of embodiment II-98, wherein the heterologous Sto7d polypeptide and the heterologous HMGN1 polypeptide are situated on the N-terminal side of the R2 protein or portion thereof.

[0583] Embodiment II-100. The nucleic acid of embodiment II-99, wherein the heterologous Sto7d polypeptide is situated on the N-terminal side of the heterologous HMGN1 polypeptide.

[0584] Embodiment II-101. The nucleic acid of any one of embodiments II-98 to II-100, wherein the R2 protein or portion thereof comprises a deletion within a native N-terminal intrinsically disordered region.

[0585] Embodiment II-102. The nucleic acid of any one of embodiments II-98 to II-101, wherein the engineered driver comprises a RNA/DNA processing polypeptide.

[0586] Embodiment II-103. The nucleic acid of any one of embodiments II-102, wherein the RNA/DNA processing polypeptide comprises a heterologous RNA helicase recruitment motif polypeptide.

[0587] Embodiment II-104. The nucleic acid of embodiment II-103, wherein the heterologous RNA helicase recruitment motif polypeptide is situated on the N-terminal side of the heterologous Sto7d polypeptide and the heterologous HMGN1 polypeptide.

[0588] Embodiment II-105. The nucleic acid of embodiment II-104, wherein the heterologous RNA helicase recruitment motif polypeptide is a TUDOR polypeptide.

[0589] Embodiment II-106. The nucleic acid of embodiment II-105, wherein the engineered driver comprises an amino acid sequence set forth in SEQ ID NO: 1234, or an amino acid sequence that is at least 70% identical thereto.

[0590] Embodiment II-107. A nucleic acid encoding an engineered driver comprising an R2 protein or portion thereof, a heterologous CTIP polypeptide, and a heterologous HMGN1 polypeptide.

[0591] Embodiment II-108. The nucleic acid of embodiment II-107, wherein the heterologous CTIP polypeptide and the heterologous HMGN1 polypeptide are situated on the N-terminal side of the R2 protein or portion thereof.

[0592] Embodiment II-109. The nucleic acid of embodiment II-108, wherein the heterologous CTIP polypeptide is situated on the N-terminal side of the heterologous HMGN1 polypeptide.

[0593] Embodiment II-110. The nucleic acid of any one of embodiments II-107 to II-109, wherein the R2 protein or portion thereof comprises a deletion in a N-terminal intrinsically disordered region or portion thereof.

[0594] Embodiment II-111. The nucleic acid of any one of embodiments II-107 to II-110, wherein the engineered driver comprises a heterologous intrinsically disordered region.

[0595] Embodiment II-112. The nucleic acid of embodiment II-111, wherein the heterologous intrinsically disordered region is situated on the N-terminal side of the heterologous CTIP polypeptide and the heterologous HMGN1 polypeptide.

[0596] Embodiment II-113. The nucleic acid of embodiment II-112, wherein the engineered driver comprises, in a N-terminal to C-terminal direction, the heterologous intrinsically disordered region, the heterologous CTIP polypeptide, the heterologous HMGN1 polypeptide, and the R2 protein or portion thereof.

[0597] Embodiment II-114. The nucleic acid of embodiment II-113, wherein the engineered driver comprises an amino acid sequence set forth in SEQ ID NOS: 679-714, or an amino acid sequence that is at least 70% identical thereto.

[0598] Embodiment II-115. The nucleic acid of any one of embodiments II-107 to II-110, wherein the engineered driver further comprises a heterologous RNA helicase polypeptide.

[0599] Embodiment II-116. The nucleic acid of embodiment II-115, wherein the engineered driver comprises, in a N-terminal to C-terminal direction, the heterologous RNA helicase polypeptide, the heterologous CTIP polypeptide, the heterologous HMGN1 polypeptide, and the R2 protein or portion thereof.

[0600] Embodiment II-117. The nucleic acid of embodiment II-116, wherein the engineered driver comprises an amino acid sequence set forth in SEQ ID NO: 783, or an amino acid sequence that is at least 70% identical thereto.

[0601] Embodiment II-118. A nucleic acid encoding an engineered driver comprising an R2 protein or portion thereof, a heterologous Sto7d polypeptide, and a heterologous HMGN1 polypeptide.

[0602] Embodiment II-119. The nucleic acid of embodiment II-118, wherein the heterologous Sto7d polypeptide and the heterologous HMGN1 polypeptide are situated on the N-terminal side of the R2 protein or portion thereof.

[0603] Embodiment II-120. The nucleic acid of embodiment II-119, wherein the heterologous Sto7d polypeptide is situated on the N-terminal side of the heterologous HMGN1 polypeptide.

[0604] Embodiment II-121. The nucleic acid of any one of embodiments II-118 to II-120, wherein the R2 protein or portion thereof comprises a deletion in a N-terminal intrinsically disordered region or portion thereof.

[0605] Embodiment II-122. The nucleic acid of any one of embodiments II-118 to II-121, comprising a heterologous TUDOR polypeptide.

[0606] Embodiment II-123. The nucleic acid of embodiment II-122, wherein the engineered driver comprises, in a N-terminal to C-terminal direction, the heterologous Tudor polypeptide, the heterologous CTIP polypeptide, the heterologous HMGN1 polypeptide, and the R2 protein or portion thereof.

[0607] Embodiment II-124. The nucleic acid of embodiment II-123, wherein the engineered driver comprises an amino acid sequence set forth in SEQ ID NO: 1234, or an amino acid sequence that is at least 70% identical thereto.

[0608] Embodiment II-125. A nucleic acid encoding an engineered driver comprising an amino acid sequence with one or more modifications relative to any one of SEQ ID NOS: 30001-30060.

[0609] Embodiment II-126. The nucleic acid of embodiment II-125, wherein the engineered driver comprises an amino acid sequence set forth in SEQ ID NOS: 30001-30060, or an amino acid sequence that is at least 70% identical thereto.

[0610] Embodiment II-127. The nucleic acid of embodiment II-125 or II-126, wherein the engineered driver comprises a reverse transcriptase domain and an endonuclease domain.

[0611] Embodiment II-128. The nucleic acid of any one of embodiments II-125 to II-127, wherein the one or more modifications is at least one amino acid substitution, at least one amino acid deletion, at least one amino acid addition, or a combination thereof.

[0612] Embodiment II-129. The nucleic acid of any one of embodiments II-125 to II-127, wherein the engineered driver comprises an amino acid sequence with one or more modifications relative to a sequence selected from the group consisting of SEQ ID NOS: 30015, 30019, 30020, 30027, and 30029.

[0613] Embodiment II-130. A nucleic acid encoding an engineered driver comprising a R2 protein or portion thereof, wherein the R2 protein or portion thereof comprises the amino acid substitution of D555K relative to SEQ ID NO:1.

[0614] Embodiment II-131. The nucleic acid of embodiment II-130, wherein the R2 protein or portion thereof comprises the amino acid substitutions P280K, D555K, T1011S, I1219N, and A1282G relative to SEQ ID NO:1.

[0615] Embodiment II-132. The nucleic acid of embodiment II-130, wherein the R2 protein or portion thereof comprises the amino acid substitutions G257R, P280K, N363H, E399M, V539G, H423Q, A442P, D537Y, K547R, D555K, R669L, N889G, F973E, T1011A, Q1019R, D1133P, K1085R, R1177G, Y1242K, and A1282G relative to SEQ ID NO:1.

[0616] Embodiment II-133. A nucleic acid encoding an engineered driver comprising an RTE polypeptide fused to a heterologous intrinsically disordered polypeptide.

[0617] Embodiment II-134. The nucleic acid of embodiment II-133, wherein the heterologous intrinsically disordered domain polypeptide comprises an amino acid sequence set forth in SEQ ID NO: 1171-1216 or an amino acid sequence that is at least 70% identical thereto.

[0618] Embodiment II-135. The nucleic acid of embodiment II-133, wherein the RTE polypeptide is lacking a native intrinsically disordered region or portion thereof.

[0619] Embodiment II-136. A nucleic acid encoding an engineered driver comprising an RTE polypeptide fused to a RNA helicase polypeptide.

[0620] Embodiment II-137. The nucleic acid of embodiment II-136, wherein the RNA helicase polypeptide is a EIF4A polypeptide.

[0621] Embodiment II-138. The nucleic acid of embodiment II-137, wherein the EIF4A polypeptide comprises an amino acid sequence set forth in SEQ ID NO: 585, or an amino acid sequence that is at least 70% identical thereto.

[0622] Embodiment II-139. The nucleic acid of embodiment II-136, wherein the RNA helicase polypeptide is a Dengue virus RNA helicase polypeptide.

[0623] Embodiment II-140. The nucleic acid of embodiment II-139, wherein the Dengue virus RNA helicase polypeptide comprises an amino acid sequence set forth in SEQ ID NO: 1875, or an amino acid sequence that is at least 70% identical thereto.

[0624] Embodiment II-141. An engineered driver encoded by the nucleic acid of any one of embodiments II-1 to II-140.

[0625] Embodiment II-142. A transgene expression system comprising:

a. a driver construct comprising the nucleic acid of any one embodiments II-1 to II-140, or the engineered driver of embodiment II-141; and

b. a gene delivery construct comprising a transgene and at least one retroelement untranslated region (RTE-UTR) derived from a RTE, wherein the at least one RTE-UTR is capable of binding the engineered driver, and the engineered driver is configured to mediate integration of the transgene into a genome.

[0626] Embodiment II-143. A transgene expression system comprising:

a. a driver construct comprising the nucleic acid of any one embodiments II-1 to II-132, or the engineered driver encoded by the nucleic acid of any one of embodiment II-1 to II-132; and

b. a gene delivery construct comprising a transgene and at least one retroelement untranslated region (RTE-UTR) derived from a R2 RTE, wherein the at least one RTE-UTR is capable of binding the engineered driver, and the engineered driver protein is configured to mediate integration of the transgene into a genome.

[0627] Embodiment II-144. The system of embodiment II-143, wherein the driver construct and the gene delivery construct are in a trans configuration and provided as separate nucleic acid constructs.

[0628] Embodiment II-145. The system of embodiment II-143, wherein the driver construct and the gene delivery construct are in a cis configuration and provided in the same nucleic acid construct.

[0629] Embodiment II-146. The system of any one of embodiments II-143 to II-145, wherein the at least one RTE-UTR comprises a 3' RTE-UTR, a 5' RTE-UTR, or a combination thereof.

[0630] Embodiment II-147. The system of embodiment II-146, wherein the 3' RTE-UTR has a nucleic acid sequence set forth in any one of SEQ ID NOS: 215, 217, 219, 20754, and 33001-33060.

[0631] Embodiment II-148. The system of embodiment II-146, wherein the 5' RTE-UTR has a nucleic acid sequence set forth in any one of SEQ ID NOS: 216, 218, 220, 20755, and 35001-35060.

[0632] Embodiment II-149. The system of embodiment II-146, wherein the 3' RTE-UTR has a nucleic acid sequence set forth in any one of SEQ ID NOS: 1338-1340, 1344-1347, 1352-1354, 1843, and 1865.

[0633] Embodiment II-150. The system of embodiment II-146 or II-147 or II-149, wherein the 5' RTE-UTR has a nucleic acid sequence set forth in any one of SEQ ID NOS: 1341-1343, 1348-1351, 1355-1360-1365, 1366, 1842, and 1866.

[0634] Embodiment II-151. The system of embodiment II-146 or II-147 or II-149, wherein the 5' RTE-UTR has a nucleic acid sequence set forth in any one of SEQ ID NOS: 1838-1850, 1873-1874, 1862-1863, 1867-1872.

[0635] Embodiment II-152. The system of any one of embodiments II-142 to II-151, wherein the R2 protein or portion thereof encoded by the driver construct is derived from the same R2 RTE as the at least one RTE-UTR comprised in the gene delivery construct.

[0636] Embodiment II-153. The system of any one of embodiments II-142 to II-151, wherein the R2 protein or portion thereof encoded by the driver construct is derived from a heterologous R2 RTE with respect to the at least one RTE-UTE comprised in the gene delivery construct.

[0637] Embodiment II-154. The system of any one of embodiments II-146 to II-148, wherein the 3' RTE-UTR has a nucleic acid sequence set forth in SEQ ID NO: 33027, and the 5' RTE-UTR has a sequence set forth in SEQ ID NO: 35027

[0638] Embodiment II-155. The system of any one of embodiments II-146-148, wherein the 3' RTE-UTR has a nucleic acid sequence set forth in SEQ ID NO: 33019, and the 5' RTE-UTR has a sequence set forth in SEQ ID NO: 35019

[0639] Embodiment II-156. The system of embodiment II-154 or II-155, wherein the R2 protein or portion thereof comprised in the engineered driver has an amino acid sequence set forth in SEQ ID NO: 1, or an amino acid sequence including at least one amino acid substitution, at least one amino acid deletion, at least one amino acid addition, or a combination thereof, relative to SEQ ID NO: 1.

[0640] Embodiment II-157. The system of embodiment II-156, wherein the engineered driver protein comprises an amino acid sequence set forth in SEQ ID NO: 636 or SEQ ID NO:1396, or an amino acid sequence that is at least 70% identical thereto.

[0641] Embodiment II-158. The system of any one of embodiments II-142 to II-157, wherein the transgene encodes a gene selected from the group consisting of: Factor VIII, Factor IX, Phenylalanine hydroxylase, ATP7B, alpha glucosidase, argininosuccinate synthetase, galactose-1-phosphate uridylyltransferase, and ornithine transcarbamylase.

[0642] Embodiment II-159. The system of any one of embodiments II-142 to II-158, wherein the transgene comprises an intron.

[0643] Embodiment II-160. The system of any one of embodiments II-142 to II-159, wherein the transgene is in a sense orientation.

[0644] Embodiment II-161. The system of any one of embodiments II-142 to II-159, wherein the transgene is in an antisense orientation.

[0645] Embodiment II-162. The system of any one of embodiments II-142 to II-161, wherein the gene delivery construct is a RNA molecule, a DNA molecule or a hybrid RNA/DNA molecule.

[0646] Embodiment II-163. A gene delivery construct comprising:

- a. a transgene; and
- b. at least one R2 retroelement untranslated region (RTE-UTR) derived from a R2 RTE, wherein the RTE-UTR is capable of binding a R2 protein, and wherein:

the at least one RTE-UTR comprises a 3' RTE-UTR, a 5' RTE-UTR, or a combination thereof,

the 3' RTE-UTR has a nucleic acid sequence set forth in any one of SEQ ID NOS: 1338-1340, 1344-1347, 1352-1354, 1843, 1865, and 33001-33060, and

the 5' RTE-UTR has a nucleic acid sequence set forth in any one of SEQ ID NOS: 1341-1343, 1348-1351, 1355-1360-1365, 1366, 1838-1850, 1873-1874, 1862-1863, 1866-1872, and 35001-35060

[0647] Embodiment II-164. The gene delivery construct of embodiment II-163, wherein the 3' RTE-UTR has a nucleic acid sequence set forth in SEQ ID NO: 33027, and the 5' RTE-UTR has a nucleic acid sequence set forth in SEQ ID NO: 35027.

[0648] Embodiment II-165. The gene delivery construct of embodiment II-163, wherein the 3' RTE-UTR has a nucleic acid sequence set forth in SEQ ID NO: 33019, and the 5' RTE-UTR has a nucleic acid sequence set forth in SEQ ID NO: 35019

[0649] Embodiment II-166. The gene delivery construct of embodiment II-163 or II-164, wherein the gene delivery construct is a RNA molecule, a DNA molecule or a hybrid RNA/DNA molecule.

[0650] Embodiment II-167. The gene delivery construct of any one of embodiments II-163 to II-166, wherein the transgene encodes a gene selected from the group consisting of: Factor VIII, Factor IX, Phenylalanine hydroxylase, ATP7B, alpha glucosidase, argininosuccinate synthetase, galactose-1-phosphate uridylyltransferase, and ornithine transcarbamylase.

[0651] Embodiment II-168. The gene delivery construct of any one of embodiments II-163 to II-167, wherein the transgene comprises an intron.

[0652] Embodiment II-169. The gene delivery construct of any one of embodiments II-163 to II-168, wherein the transgene is in a sense orientation.

[0653] Embodiment II-170. The gene delivery construct of any one of embodiments II-163 to II-169, wherein the transgene is in an antisense orientation.

[0654] Embodiment II-171. The gene delivery construct of any one of embodiments II-163 to II-170, wherein the gene delivery construct is a RNA molecule, a DNA molecule or a hybrid RNA/DNA molecule.

[0655] Embodiment II-172. The gene delivery construct of any one of embodiments II-163 to II-166, wherein the transgene encodes a gene selected from the group consisting of: Factor VIII, Factor IX, Phenylalanine hydroxylase, ATP7B, alpha glucosidase, argininosuccinate synthetase, galactose-1-phosphate uridylyltransferase, and ornithine transcarbamylase.

[0656] Embodiment II-173. A lipid nanoparticle (LNP) comprising:

- a. the nucleic acid of any one of embodiments II-1 to II-140;
- b. the engineered driver of embodiment II-141;
- c. the transgene expression system of any one of embodiments II-142 to II-162; or
- d. the gene delivery construct of any one of embodiment II-163 to II-172; or
- e. a combination thereof.

[0657] Embodiment II-174. The LNP of embodiment II-173, wherein the LNP comprises an ionizable lipid, a PEGylated lipid, an amino lipid, a phospholipid, a cholesterol, a sterol, a non-cationic lipid, or a combination thereof.

[0658] Embodiment II-175. A pharmaceutical composition comprising:

- a. the nucleic acid of any one of embodiments II-1 to II-140;
- b. the engineered driver of embodiment II-141;
- c. the transgene expression system of any one of embodiments II-142 to II-162;
- d. the gene delivery construct of any one of embodiment II-163 to II-172;
- e. the LNP of any one of embodiments II-173 to II-174; or
- f. a combination thereof;

and a pharmaceutically acceptable carrier.

[0659] Embodiment II-176. A method of promoting integration of a transgene into a genome of a cell, the method comprising contacting a cell with:

- a. the nucleic acid of any one of embodiments II-1 to II-140;
- b. the engineered driver of embodiment II-141;
- c. the transgene expression system of any one of embodiments II-142 to II-162;
- d. the gene delivery construct of any one of embodiment II-163 to II-172;
- e. the LNP of any one of embodiments II-173 to II-174;
- f. a composition of embodiment II-175; or
- g. a combination thereof.

[0660] Embodiment II-177. The method of embodiment II-176, wherein the cell is a quiescent cell.

[0661] Embodiment II-178. The method of embodiment II-176, wherein the cell is an actively dividing cell.

[0662] Embodiment II-179. The method of embodiment II-176, wherein the cell is T-cell.

[0663] Embodiment II-180. The method of embodiment II-176, wherein the cell is a cancer cell.

[0664] Embodiment II-181. The method of any one of embodiments II-176 to II-177 and II-180, wherein the cell is an endothelial cell, an astrocyte, an osteoblast, a lung cell, a neuron, a hepatocyte, a pancreatic cell, a myocyte, a keratinocyte, a bladder cell, or an adipocyte.

[0665] Embodiment II-182. The method of any one of embodiments II-176 to II-181, wherein the method takes place ex vivo.

[0666] Embodiment II-183. The method of any one of embodiments II-176 to II-181, wherein the cell is an allogeneic cell.

[0667] Embodiment II-184. The method of any one of embodiments II-176 to II-181, wherein the cell is an autologous cell.

[0668] Embodiment II-185. The method of any one of embodiments II-176 to II-184, wherein the method takes place *ex vivo*.

[0669] Embodiment II-186. The method of any one of embodiments II-176 to II-184, wherein the method takes place *in vivo*.

[0670] Embodiment II-187. A population of cells prepared by the method of any one of embodiments II-176 to II-186.

[0671] Embodiment II-188. A method of treating a disease in a subject in need thereof, comprising administering to the subject a therapeutically effective amount of:

- a. the nucleic acid of any one of embodiments II-1 to II-140;
- b. the engineered driver of embodiment II-141;
- c. the transgene expression system of any one of embodiments II-142 to II-162;
- d. the gene delivery construct of any one of embodiment II-163 to II-172;
- e. the LNP of any one of embodiments II-173 to II-174;
- f. the composition of embodiment II-175; or
- g. a combination thereof.

[0672] Embodiment II-189. The nucleic acid of any one of embodiments II-1 to II-140, the engineered driver of embodiment II-141, the transgene expression system of any one of embodiments II-142 to II-162, the gene delivery construct of any one of embodiment II-163 to II-172, the LNP of any one of embodiments II-173 to II-174, the composition of embodiment II-175, or a combination thereof, for use in a method of treating the human or animal body by therapy.

[0673] Embodiment II-190. Use of the nucleic acid of any one of embodiments II-1 to II-140, the engineered driver of embodiment II-141, the transgene expression system of any one of

embodiments II-142 to II-162, the gene delivery construct of any one of embodiment II-163 to II-172, the LNP of any one of embodiments II-173 to II-174, the composition of embodiment II-175, or a combination thereof, for the manufacture of a medicament.

[0674] The inventions are further illustrated by the following examples. The examples are provided for illustrative purposes only and are not to be construed as limiting the scope or content of the inventions in any way.

EXAMPLES

Example 1: Engineered drivers and fusion proteins

[0675] Non-limiting examples of engineered drivers disclosed herein below and in subsequent examples that can be used (e.g., directly or encoded on an RNA and/or DNA molecule) to promote insertion of a transgene into a target nucleic acid (e.g., a host genome) are provided. One or more of the engineered drivers herein below and in subsequent examples can be provided alone or in combination with other proteins or engineered proteins. Accordingly, one or more of the following heterologous polypeptides that are provided herein below and in subsequent examples as N-terminal fusions also could be provided as C-terminal fusions and/or as internal fusions. Accordingly, one or more of the following heterologous polypeptides that are provided herein below and in subsequent examples as C-terminal fusions also could be provided as N-terminal fusions and/or as internal fusions. In some embodiments, an engineered driver includes a N-terminal fusion of one or more (e.g., 1, 2, 3, or more) heterologous polypeptides, a C-terminal fusion of one or more (e.g., 1, 2, 3, or more) heterologous polypeptides and/or an internal fusion of one or more (e.g., 1, 2, 3, or more) heterologous polypeptides.

[0676] Provided hereinbelow, in this and later examples, are non-limiting examples of engineered driver comprising an R2 RTE polypeptide (e.g., a R2 protein or portion thereof) of any one of R2-TG, R2-ZA, or R4-1-PH RTEs, whose WT R2 proteins are set forth in SEQ ID NOs: 1, 2 and 3 respectively, according to the description herein. It will be appreciated that SEQ ID NOs: 1, 2, and 3 may be replaced with other, or modified, RTEs as disclosed throughout the specification.

[0677] SEQ ID NO: 120 sets forth the amino acid sequence of a fusion protein comprising an internal RAD51 within R2-TG.

[0678] SEQ ID NO: 121 sets forth the amino acid sequence of a fusion protein comprising an internal UL12 R2-TG fusion. SEQ ID NO: 321 sets forth the corresponding DNA sequence.

[0679] SEQ ID NO: 122 sets forth the amino acid sequence of a fusion protein comprising an internal BRCA2-derived peptide R2-TG fusion. SEQ ID NO: 322 sets forth the corresponding DNA sequence.

[0680] SEQ ID NO: 123 sets forth the amino acid sequence of a fusion protein comprising an N-terminal HMGN1 R2-TG fusion combined with a C-terminal HMGB1 fusion. SEQ ID NO: 323 sets forth the corresponding DNA sequence.

[0681] SEQ ID NO: 124 sets forth the amino acid sequence of a fusion protein comprising a C-terminal HMGB1 polypeptide R4-AC fusion.

[0682] SEQ ID NO: 125 sets forth the amino acid sequence of a fusion protein comprising N-terminal HMGN1 polypeptide and UL12 polypeptide and HMGB1 polypeptide fused at the C-terminus of the R2-TG polypeptide. SEQ ID NO: 325 sets forth the corresponding DNA sequence.

[0683] SEQ ID NO: 126 sets forth the amino acid sequence of a fusion protein comprising a C-terminal Sto7d R2-TG fusion. SEQ ID NO: 326 sets forth the corresponding DNA sequence.

[0684] SEQ ID NO: 127 sets forth the amino acid sequence of a fusion protein comprising a C-terminal UL12 R2-TG fusion.

[0685] SEQ ID NO: 128 sets forth the amino acid sequence of a fusion protein comprising a C-terminal sSo7d R2-TG fusion.

[0686] SEQ ID NO: 129 sets forth the amino acid sequence of a fusion protein comprising a C-terminal StkC R2-TG fusion.

[0687] SEQ ID NO: 130 sets forth the amino acid sequence of a fusion protein comprising a sequence FEN1 PCNA recruitment peptide R2-TG fusion.

[0688] SEQ ID NO: 131 sets forth the amino acid sequence of a fusion protein comprising a P21 PCNA recruitment peptide to replace the native PCNA motif in R2-TG.

[0689] SEQ ID NO: 132 sets forth the amino acid sequence of a fusion protein comprising a C-terminal RPA3 R4-AC fusion.

[0690] SEQ ID NO: 133 sets forth the amino acid sequence of a fusion protein comprising a RTE chimera between different R2 RTEs (R2-ZA and R2-TG).

[0691] SEQ ID NO: 134 sets forth the amino acid sequence of a fusion protein comprising a RTE chimera between different R2 RTE polypeptides.

[0692] SEQ ID NO: 135 sets forth the amino acid sequence of a fusion protein comprising an N-terminal SV40 NLS R2-TG fusion.

[0693] SEQ ID NO: 136 sets forth the amino acid sequence of a fusion protein comprising a C-terminal RAD6 fusion R2-TG fusion.

[0694] SEQ ID NO: 137 sets forth the amino acid sequence of a fusion protein comprising an N-terminal MDM2 p53 inhibitory peptide R2-TG fusion. SEQ ID NO: 337 sets forth the corresponding DNA sequence.

[0695] SEQ ID NO: 138 sets forth the amino acid sequence of a fusion protein comprising a C-terminal TOPBP1 peptide R2-TG fusion.

[0696] SEQ ID NO: 139 sets forth the amino acid sequence of a fusion protein comprising a C-terminal PARP1 peptide R2-TG fusion.

[0697] SEQ ID NO: 140 sets forth the amino acid sequence of a fusion protein comprising a C-terminal ANKRD28 peptide R2-TG fusion.

[0698] SEQ ID NO: 141 sets forth the amino acid sequence of a fusion protein comprising a C-terminal MDC1 peptide R2-TG fusion.

[0699] SEQ ID NO: 142 sets forth the amino acid sequence of a fusion protein comprising a C-terminal MSH4 peptide R2-TG fusion.

[0700] SEQ ID NO: 143 sets forth the amino acid sequence of a fusion protein comprising a C-terminal RAD17 peptide R2-TG fusion.

[0701] SEQ ID NO: 144 sets forth the amino acid sequence of a fusion protein comprising a C-terminal SCML1 peptide R2-TG fusion.

[0702] SEQ ID NO: 145 sets forth the amino acid sequence of a fusion protein comprising a C-terminal CDKN2A peptide R2-TG fusion.

[0703] SEQ ID NO: 146 sets forth the amino acid sequence of a fusion protein comprising a C-terminal CHAF1A PCNA recruitment peptide R2-TG fusion.

[0704] SEQ ID NO: 147 sets forth the amino acid sequence of a fusion protein comprising an N-terminal C. Stercoris RecT (csRecT) peptide R2-TG fusion.

[0705] SEQ ID NO: 148 sets forth the amino acid sequence of a fusion protein comprising an N-terminal P. Aeruginosa RecT (paRecT) peptide R2-TG fusion.

[0706] SEQ ID NO: 481 sets forth the amino acid sequence of a fusion protein comprising an N-terminal P. Aeruginosa RecT (paRecT) peptide R2-TG fusion using a modified R2-TG protein sequence with 3 amino acid substitutions and an alternative start site.

[0707] SEQ ID NO: 149 sets forth the amino acid sequence of a fusion protein comprising an N-terminal zinc finger domain targeting AAVS1 safe harbor R2-TG fusion.

[0708] SEQ ID NO: 150 sets forth the amino acid sequence of an R2-TG without the EN domain. In this protein the R2-TG EN domain was removed.

[0709] SEQ ID NO: 151 sets forth the amino acid sequence of a fusion protein comprising a nickase Cas9 R2-TG fusion.

[0710] SEQ ID NO: 152 sets forth the amino acid sequence of a fusion protein comprising a dead Cas9 R2-TG fusion. In this fusion the first 470AA (DNA binding regions) of R2-TG were removed

[0711] SEQ ID NO: 153 sets forth the amino acid sequence of a fusion protein comprising a nickase Cas9 R2-TG fusion.

[0712] SEQ ID NO: 154 sets forth the amino acid sequence of a fusion protein comprising a nickase Cas9 R2-TG fusion. In this fusion the R2-TG EN domain (C-terminal 163 amino acids) was removed.

[0713] SEQ ID NO: 155 sets forth the amino acid sequence of a fusion protein comprising a dead Cas9 R2-TG fusion.

[0714] SEQ ID NO: 156 sets forth the amino acid sequence of a fusion protein comprising a Cas9 nickase fused to an endonuclease-deleted R2-TG.

[0715] Nucleic acids (e.g., RNA or DNA) encoding one or more of these engineered driver, or the engineered drivers provided in the subsequent examples herein below, can be provided *in cis* or *in trans* along with a gene delivery construct a transgene (e.g., flanked by terminal sequences).

Example 2: Methods for evaluation of transgene integration

[0716] An integration assay was performed to evaluate the percentage of cells in which a stable reporter (e.g., GFP) construct was integrated into genomic DNA by a driver (e.g., an engineered driver of the disclosure). A GFP reporter construct was evaluated in combination with different engineered driver constructs in trans. The reporter construct comprises (unless otherwise indicated) an antisense expression cassette for GFP (driven by a CMV promoter and containing a polyadenylation signal from the thymidine kinase gene from the herpes simplex virus) and a 3' RTE-UTR of a R2 RTE with 20 nucleotide homology arms to the 28S rDNA integration site. The trans driver construct contains the single ORF encoding a driver (e.g. an engineered driver of the disclosure or a WT R2 protein used in certain cases as a control) and a polyadenylated SV40 sequence ending in A30-N10-A70 (wherein the N10 is a non-polyA sequence). The reporter and driver constructs were configured with T7 promoters and a unique Type IIS restriction site at the 3' end. Upon restriction, these DNAs were used in in vitro transcription (IVT) reactions using NEB HiScribe™ T7 High Yield RNA Synthesis Kit to produce the constructs in RNA form.

[0717] Briefly, the plasmid was first dissolved and let stand at room temperature. Then, to assess the concentration of the plasmid, a nanodrop was used to measure the absorbance of the sample. The plasmid was then restricted with a restriction enzyme and cleaned using AMPure

XP beads mixed into the solution. Using a magnetic tube rack the solution was aspirated with 70% ethanol added and incubated at room temperature three times. The ethanol was then removed, and the beads dried. The plasmids were then resuspended using an elution buffer, dried and resuspended in water. Once resuspended the plasmid concentration was measured.

[0718] For *in vitro* transcription (IVT) production the NEB HiScribe™ T7 High Yield RNA Synthesis Kit was used followed by the TURBO™ DNase kit AM2238. The RNA transcript was purified using Monarch® RNA cleanup kit (NE Biolab).

[0719] To check the quality of the sample, quantification was done using Nanodrop™ (ThermoFisher). Using the Agilent TapeStation device and Controller software the quality and uniformity were measured.

[0720] Transgene integration assays were conducted in U2OS cells (unless otherwise indicated). Cells were seeded 24 hours prior to transfection. Transfection was done using Lipofectamine™ MessengerMAX™ (unless otherwise indicated). The mRNA master mix was added to the diluted MessengerMAX™ reagent and incubated. The new RNA-lipid complex was added to the cells and incubated overnight. Reporter integration was checked by measuring the percent of GFP expressing cells after 24 hours using FACS analyses and fluorescent microscopy.

[0721] Integration was assessed based on the percentage of GFP positive cells (% GFP positive cells) after 24h, with a higher percentage of GFP positive cells being indicative of higher efficiency of integration. % GFP positive cells was assessed by FACS following 24 hours after transfection with RNA. Results shown here are an average of at least two biological replicates.

Example 3: Engineered R2 RTE proteins

[0722] Nucleic acids were designed and produced to encode non-limiting examples of engineered R2-TG protein with one or more point mutations in the RNA binding domain and the reverse transcriptase domain. These were tested and evaluated using an experimental transgene integration assay described in Example 2. IVT of the different RTE drivers and reporters in cis configuration was carried out as described above to produce the RNA versions of the driver and reporter constructs. U2OS cells were used in 24-well plates, at 100K cells/well. 1000ng RNA (driver and reporter constructs) was transfected with 1.2uL Lipofectamine™.

[0723] Non-limiting examples of the heterologous polypeptides include the following mutations relative to the R2-TG wild type polypeptide having an amino acid sequence set forth in SEQ ID NO:1 are L18P, E19L, S26N, V27R, A29T, I30A, R32G, N35I, S36R, L37K, A38R, L49F, G209R, K289R, L316H, E399M, E472G, Y477H, T514R, D537Y, T538K, H544Q, D555K, N562D, W762R, R822K, S895K, Q909N, R964K, G966N, T989Y, I1003M, T1011S, A1012G, L1013R, T1016L, Q1019R, E1027D, P1033R, C1079K, K1085R, L1089R, H1090Y, I1219N, A1282G, T989I, K579R, K547R, V660I, T608P, F956I, I1003L, T1016S, T680S, L594P, P52R, P418S, P78G, L51G, P78V, P180S, V539G, P180K, P437R, V539R, N640A, W677C, V687W, R866I, A747G, W762E, T908I, W959K, D1037N, Y1194L, R1292P, K1210R, K993R, K755R, C518R, K314R, K1182R, K385M, P437M, T448Y, C518Y, E1008R, T305R, P280K, A1012R, N269R, L213K, E214K, Q215K, E214R, T1011R, E1015R, L213R, G257M, G209F, W762D, Y1194L, V687W, W677C, R1292P, T908I, W959K, N640A, Q909N, R866I, A1282G, W762E, W959K, T908I, Y579R, T1011A, Y571H, E630D, D1075E, L1089M, A442P, H423Q, E173D, P188L, C145G, N301G, R355K, G257R, V251G, T603A, N363H, P1033S, F667Y, F1365L, V501L, G257R+C145G, W762R+C145G, G257R+E173D, N363H+E173D, A442P+E173D, W762R+E173D, G257R+P188L, W762R+P188L, G257R+V251G, S895K+V251G, N363H+G257R, R568S, E542K, P985R, G561R, Y563R, L1184K, H1030R, A557R, I558R, Q567R, N271W, E350R, L316R, F304K, D925K, A317W, E1010R, Q1021R, E419R, T1011F, E769R, W345R, V110R, R568G, F135G, N104Y, V142Y, P115K, A111Y, L129R, V98R, D126R, E144R, Q112R, Y136R, F148K, T1016K, K700D.

Results:

[0724] FIGS. 4A and 4B show results of integration assays using drivers comprising heterologous polypeptide fusions with, and mutations in, R2-TG protein (furthest left bar in 4A). In this experiment, different retrotransposable element driver constructs were used with the same reporter construct (GFP). The drivers were mutated R2-TG protein constructs described above. The driver constructs were as follows, from left to right: R2-TG wildtype control, N-terminal fusion UL12 peptide, C-terminal fusion BRCA2 polypeptide, C-terminal fusion Sto7D polypeptide, N-terminal fusion Sso7D polypeptide, N-terminal fusion StkC polypeptide, and N-terminal fusion FEN1-PCNA polypeptide, and the following amino acid substitutions (with respect to SEQ ID NO: 1), E399M, Y477H, D537Y, D555K, W762R, R822K, S895K, I1003M,

T1011S, A1012G, L1013R, T1016C, Q1019R, E1027D, P1033R, C1079K, K1085R, H1090Y, I1219N, A1282G, T981I.

[0725] FIG. 4B. shows results from driver constructs encoding R2-TG protein with the following amino acid substitutions(with respect to SEQ ID NO: 1), from left to right: K579R, K547R, T608P, P956I, I1003L, T680S, P78G, L51G, P78V, L51G, P78V, V539G, P180K, V539R, W677C, A747G, K314R, K1182R, K385M, P437M, T448Y, C518Y, T305R, P280K, N269R, L213K, E214K, Q215K, E214R, T1011R.

[0726] Various mutations (amino acid substitutions) and mutation combinations showed enhanced activity relative to the wild-type construct. For example, as compared to R2-TG wild type (SEQ ID NO:1) engineered polypeptides which included at least one of the following amino acid substitutions each individually improved activity by at least 10%: E399M, Y477H, D555K, W762R, R822K, H1090Y, F956I, P78V, W677C, K314R, K385M, P437M, T448Y, C518Y, T305R, P280K, N269R, L213K, E214K, Q215K, E214R, T1011R. N-terminal fusion of Sto7D (SEQ ID NO: 42) to R2-TG also enhanced activity by at least 10% relative to the reference R2-TG retroelement polypeptide.

[0727] Altogether, these results illustrate several important aspects disclosed herein. In some embodiments, one or more amino acid substitutions in the reverse transcription domain, for example those that promote RT interactions with RNA and/or DNA (e.g., including one or more amino acid substitutions such as P280K and D555K in R2-TG) significantly enhance transgene integration. Significantly enhance refers to at least a 10% (and to over 50%) improvement in %GFP positive cells compared to wildtype retroelement. Secondly, in some embodiments, fusion of DNA-binding proteins (For example Sto7D) may improve transgene integration.

[0728] SEQ ID NOs: 230-238 set forth amino acid sequences of R2-TG protein comprising amino acid substitutions shown to improve at least one of efficiency, specificity, accuracy, fidelity or processivity compared to the polypeptide set forth in SEQ ID NO:1

[0729] SEQ ID NO: 230 sets forth the mutated amino acid sequence of R2-TG protein with the D555K amino acid substitutions.

[0730] SEQ ID NO: 231 sets forth the mutated amino acid sequence of R2-TG protein with the P280 and K579R amino acid substitutions.

[0731] SEQ ID NO: 232 sets forth the mutated amino acid sequence of a R2-TG protein with the A1012G and D555K amino acid substitutions.

[0732] SEQ ID NO: 233 sets forth the mutated amino acid sequence of a R2-TG protein with the C145G amino acid substitutions.

[0733] SEQ ID NO: 234 sets forth the mutated amino acid sequence of an R2-TG protein with the T1011S, I1219N, D555K and P280K amino acid substitutions.

[0734] SEQ ID NO: 235 sets forth the mutated amino acid sequence of an R2-TG protein with the G257R and C145G amino acid substitution.

[0735] SEQ ID NO: 236 sets forth the mutated amino acid sequence of a R2-TG protein with the F667G amino acid substitution.

[0736] SEQ ID NO: 237 sets forth the mutated amino acid sequence of a R2-TG protein with the K700D amino acid substitution.

[0737] SEQ ID NO: 238 sets forth the mutated amino acid sequence of a R2-TG protein with the F148K amino acid substitution.

N-terminal R2-TG deletion and protein domain fusions

[0738] Nucleic acids were designed and produced to encode non-limiting examples of engineered proteins comprising an R2-TG protein with N-terminal deletion (e.g., of the native intrinsically disordered region or portions thereof) and heterologous polypeptide fusions. These were tested and evaluated using an experimental transgene integration assay described in Example 2. Non-limiting examples of the heterologous polypeptides include the following modifications: N-terminal V161 truncation fused to N-terminal HMGN1 polypeptide, N-terminal V161 truncation fused to N-terminal Sto7D polypeptide, N-terminal V161 truncation fused to N-terminal Sso7D polypeptide, N-terminal V161 truncation fused to N-terminal P21 polypeptide, N-terminal V161 truncation fused to N-terminal FEN1 polypeptide, N-terminal V161 truncation fused to N-terminal UL12 polypeptide, N-terminal V161 truncation fused to N-terminal BRCA2 polypeptide, N-terminal V161 truncation fused to N-terminal RAD51 polypeptide, N-terminal V161 truncation fused to N-terminal CTIP polypeptide, N-terminal E184 truncation fused to N-terminal HMGN1 polypeptide, N-terminal E184 truncation fused to

N-terminal Sto7D polypeptide, N-terminal E184 truncation fused to N-terminal Sso7D polypeptide, N-terminal E184 truncation fused to N-terminal P21 polypeptide, N-terminal E184 truncation fused to N-terminal FEN1 polypeptide, N-terminal E184 truncation fused to N-terminal UL12 polypeptide, N-terminal E184 truncation fused to N-terminal BRCA2 polypeptide, N-terminal E184 truncation fused to N-terminal RAD51 polypeptide, and N-terminal E184 truncation fused to N-terminal CTIP polypeptide.

[0739] Materials and methods are described in Example 2, above. Results shown here are an average of at least two biological replicates.

[0740] FIG. 4C is a bar graph showing results of integration assays using drivers comprising peptide fusions and deletions (truncation 5' to amino acid V161 or E184) in the R2-TG protein having amino acid sequences set forth in SEQ ID NOs: 250-267. In this experiment, different engineered driver constructs were used with the GFP reporter construct. The drivers were mutated engineered R2-TG drivers as described above. The results shown are as follows, from left to right: GFP reporter negative control, wildtype RT-TG control, N-terminal V161 truncation fused to N-terminal HMGN1 polypeptide, N-terminal V161 truncation fused to N-terminal Sto7D polypeptide, N-terminal V161 truncation fused to N-terminal P21 polypeptide, N-terminal V161 truncation fused to N-terminal FEN1 polypeptide, N-terminal V161 truncation fused to N-terminal CTIP polypeptide, N-terminal E184 truncation fused to N-terminal Sso7D polypeptide, N-terminal E184 truncation fused to N-terminal P21 polypeptide, N-terminal E184 truncation fused to N-terminal RAD51 polypeptide, N-terminal V161 truncation fused to N-terminal UL12 polypeptide, V161 truncated control, E184 truncated control, and positive control. The driver constructs are described, hereinbelow:

[0741] **SEQ ID NO: 250** sets forth the amino acid sequence of a R2-TG N-terminal truncation (up to residue 161) with N-terminal fusion of HMGN1 polypeptide. The HMGN1 polypeptide is fused at the C-terminus via a linker to an N-terminal R2-TG (V161) protein.

[0742] **SEQ ID NO: 251** sets forth the amino acid sequence of a R2-TG N-terminal truncation (up to residue 161) with N-terminal fusion of Sto7D polypeptide. The Sto7D polypeptide is fused at the C-terminus via a linker to an N-terminal R2-TG (V161) protein.

[0743] SEQ ID NO: 252 sets forth the amino acid sequence of a R2-TG N-terminal truncation (up to residue 161) with N-terminal fusion of Sso7D polypeptide. The Sso7D polypeptide is fused at the C-terminus via a linker to an N-terminal R2-TG (V161) protein.

[0744] SEQ ID NO: 253 sets forth the amino acid sequence of a R2-TG N-terminal truncation (up to residue 161) with N-terminal fusion of P21-PCNA polypeptide. The P21-PCNA polypeptide is fused at the C-terminus via a linker to an N-terminal R2-TG (V161) protein.

[0745] SEQ ID NO: 254 sets forth the amino acid sequence of a R2-TG N-terminal truncation (up to residue 161) with N-terminal fusion of FEN-PCNA polypeptide. The FEN-PCNA polypeptide is fused at the C-terminus via a linker to an N-terminal R2-TG (V161) protein.

[0746] SEQ ID NO: 255 sets forth the amino acid sequence of a R2-TG N-terminal truncation (up to residue 161) with N-terminal fusion of UL12 polypeptide. The UL12 polypeptide is fused at the C-terminus via a linker to an N-terminal R2-TG (V161) protein.

[0747] SEQ ID NO: 256 sets forth the amino acid sequence of a R2-TG N-terminal truncation (up to residue 161) with N-terminal fusion of BRCA2 polypeptide. The BRCA2 polypeptide is fused at the C-terminus via a linker to an N-terminal R2-TG (V161) protein.

[0748] SEQ ID NO: 257 sets forth the amino acid sequence of a R2-TG N-terminal truncation (up to residue 161) with N-terminal fusion of RAD51 polypeptide. The RAD51 polypeptide is fused at the C-terminus via a linker to an N-terminal R2-TG (V161) protein.

[0749] SEQ ID NO: 258 sets forth the amino acid sequence of a R2-TG N-terminal truncation (up to residue 161) with N-terminal fusion of CTIP polypeptide. The CTIP polypeptide is fused at the C-terminus via a linker to an N-terminal R2-TG (V161) protein.

[0750] SEQ ID NO: 259 sets forth the amino acid sequence of a R2-TG N-terminal truncation (up to residue 184) with N-terminal fusion of HMGN1 polypeptide. The HMGN1 polypeptide is fused at the C-terminus via a linker to an N-terminal R2-TG (E184) protein.

[0751] SEQ ID NO: 260 sets forth the amino acid sequence of a R2-TG N-terminal truncation (up to residue 184) with N-terminal fusion of Sto7D polypeptide. The Sto7D polypeptide is fused at the C-terminus via a linker to an N-terminal R2-TG (E184) protein.

[0752] SEQ ID NO: 261 sets forth the amino acid sequence of a R2-TG N-terminal truncation (up to residue 184) with N-terminal fusion of Sso7D polypeptide. The Sso7D polypeptide is fused at the C-terminus via a linker to an N-terminal R2-TG (E184) protein.

[0753] SEQ ID NO: 262 sets forth the amino acid sequence of a R2-TG N-terminal truncation (up to residue 184) with N-terminal fusion of P21-PCNA polypeptide. The P21-PCNA polypeptide is fused at the C-terminus via a linker to an N-terminal R2-TG (E184) protein.

[0754] SEQ ID NO: 263 sets forth the amino acid sequence of a R2-TG N-terminal truncation (up to residue 184) with N-terminal fusion of FEN-PCNA polypeptide. The FEN-PCNA polypeptide is fused at the C-terminus via a linker to an N-terminal R2-TG (E184) protein.

[0755] SEQ ID NO: 264 sets forth the amino acid sequence of a R2-TG N-terminal truncation (up to residue 184) with N-terminal fusion of UL12 polypeptide. The UL12 polypeptide is fused at the C-terminus via a linker to an N-terminal R2-TG (E184) protein.

[0756] SEQ ID NO: 265 sets forth the amino acid sequence of a R2-TG N-terminal truncation (up to residue 184) with N-terminal fusion of BRCA2 polypeptide. The BRCA2 polypeptide is fused at the C-terminus via a linker to an N-terminal R2-TG (E184) protein.

[0757] SEQ ID NO: 266 sets forth the amino acid sequence of a R2-TG N-terminal truncation (up to residue 184) with N-terminal fusion of RAD51 polypeptide. The RAD51 polypeptide is fused at the C-terminus via a linker to an N-terminal R2-TG (E184) protein.

[0758] SEQ ID NO: 267 sets forth the amino acid sequence of a R2-TG N-terminal truncation (up to residue 184) with N-terminal fusion of CTIP polypeptide. The CTIP polypeptide is fused at the C-terminus via a linker to an N-terminal R2-TG (E184) protein.

R2-ZA

[0759] Nucleic acids were designed and produced to encode non-limiting examples of engineered drivers comprising a R2-ZA protein with one or more amino acid substitutions in the RNA binding domain and the reverse transcriptase domain. These engineered drivers were tested and evaluated using an experimental transgene integration assay described in Example 2. Non-limiting examples of the heterologous polypeptides include the following mutations relative to an R2-ZA wild type polypeptide having an amino acid sequence set forth in SEQ ID NO: 2.

[0760] Tested mutations: P87V, A289P, M287A, C482S, G1057D, S931R, V996S, L1100F, M685L, S547P, V916I, A1256E, T654K, L477C, M1259A, T1120K, N332G, M358L, V1230L+F585L, I728V+F585L, I728V+M685L, Y919P, Y1322I, T950A, F1182Q, F585P, M376P, V1230L, F1294W, F1294W+I728V, F1294W+F585L, V1230L+M685L, F1294W+M685L, V1230L+S547P, N843G, P151K, Y1079T, V748W, V748D, L1277I, E1026T, F1272G, F1272R, D584L, M1248G, F1024A, L705C, Y1162Q, M640S, R642K, C314L, R1131G, A965S, Y1044L, D1236G, A970L, Q1039R, C474Y, K501Y, H393Y, T404Y, K493Y, S494G.

[0761] U2OS cells were used in 24-well plates, at 100K cells/well. 1000ng RNA was transfected with 1.2uL Lipofectamine™. IVT of different RTE drivers and the GFP reporter in cis configuration was carried out as described above. Integration was assessed based on the percentage of GFP positive cells (% GFP positive cells) after 24h, with a higher percentage of GFP positive cells being indicative of higher levels of integration. % GFP positive cells was assessed by FACS following 24 hours after transfection with RNA. Results shown are an average of at least two biological replicates.

[0762] FIG. 5 show results of integration assays using drivers comprising peptide fusions and mutations in R2-ZA protein relative to an amino acid sequence set forth in SEQ ID NO:2. In this experiment, different retrotransposable element driver constructs were used with the same reporter construct (GFP).

[0763] In FIG. 5, the drivers were mutated R2-ZA protein constructs described above. The amino acid substitutions are as follows, from left to right: The constructs are as follows, from left to right: A289P, M287A, G1057D, S931R, A289P, L1100F, M287A, S547, V916I, A1256E, T654K, L577C, M1259A, N332G, M358L, V1230 and F585L, I728V and F585L, I728V and M685L, Y919P, Y1322I, T950A, F1182Q, F585P, M376P, V1230L, F1294W, F1294 and I728V, F1294W and F585L, V1230L and M685L, F1294W and M685L, V1230L and S547P. Reference control (WT) is furthest right bar.

[0764] In FIG. 6 the constructs are as follows, from left to right: from left to right: P151K, Y1079T, V748W, V748D, E1026T, F1272G, F1272R, D584L, F1024A, L705C, Y1162Q, M640S, R642K, C314L, R1131G, A965S, D1236G, A970L, Q1039R, C474Y, K501Y, H393TM T404Y K493Y, S494G, P78V. Reference control (WT) is furthest right bar.

[0765] Mutations that improved the activity above reference R2-ZA SEQ ID NO:2 are G1057D, S931R, A289P, V996S, L1100F, M287A, C482S, M685L, S547P, V916I, A1256E, M1259A, N332G, M358L, Y919P, T950A, M376P, P151K, D584L, Y1162Q, M640S, C314L, R1131G, A965S, Y1044L, D1236G, A970L, Q1039R, C474Y, T404Y.

[0766] In general, and without wishing to be bound to theory, improving RNA/DNA binding interaction of the RT domain through addition of positive charge (S931R, Q1039R) may improve activity of the R2-ZA retrotransposable element. In addition, substitution of methionine residues (M287A, M358L, M376P, M640S, M685L, M1259A) may lead to improved activity by stabilizing the protein. Methionine to Leucine mutations may improve protein stability by reducing entropic cost of folding (Lipscomb et al, 1998), and render the polypeptide less sensitive to oxidation (Kim et al, 2001).

R4-AC

[0767] Nucleic acids were designed and produced to encode non-limiting examples of engineered proteins comprising an R4-AC protein with one or more amino acid substitutions in the RNA binding domain and the reverse transcriptase domain. These were tested and evaluated using an experimental transgene integration assay described in Example 2. Non-limiting examples of the heterologous polypeptides include the following amino acid substitutions relative to an R4-AC wild type polypeptide having an amino acid sequence set forth in SEQ ID NO:3:

Q6S, R71K, S166P, Q195K, R287K, W303K, L405Q, M412E, N522I, C579P, C579S, G583R, Y604S, I629L, I629E, V790D, V790M, V790A, T866A, A877T, C966G, V1029T, T1034R, R1057N, R1068K, R1103K, T581N, H1128R, L1011V, I644T, F233Y, A153E, D124N, T956A, D659Y, H740V, T889A, R266I, Q118K, T410S, T410K, H1015V, A25H, V361T, R206S, P1098S, D814S, T1038V, G1054F, I986L, S460P, F614R, N639R, K747M, K1073R, K62M, K62I, K326L, K326R, K537R, K837R, A697Q, Y52W, C795S, T518N, R416A, C719L, W988K, M406P, N522F, A324R, E331R, S257W, S257K, P69R, I104K, I104R, F102R, Q70R, G73R, E111R, E328R, I327R, S276R, T258R.

[0768] Fusion constructs comprising a R4-AC protein and heterologous polypeptide as described were also tested. FIG. 7 shows the results of testing a subset of these compositions.

[0769] FIG. 7 shows results of integration assays using drivers comprising polypeptide fusions which include mutations in R4-AC protein. In this experiment, different retrotransposable element driver constructs were tested with the GFP reporter construct. The drivers were modified R4-AC driver constructs described above. The constructs are as follows, from left to right: negative control, R4-AC wildtype (SEQ ID NO:3), C-terminal UL12 polypeptide fusion, C-terminal StkC polypeptide fusion, N-terminal Sto7D polypeptide fusion, N-terminal Sso7D polypeptide fusion, N-terminal StkC polypeptide fusion, FEN polypeptide fusion, P21 polypeptide fusion insertion site 1 (SEQ ID NO: 157), P21 polypeptide insertion site 2 (SEQ ID NO:158), C-terminal RPA polypeptide fusion, Q195K substitution, W303K substitution, M412E substitution, N522I substitution, C579P substitution, Y604S substitution, I629L substitution, V790D substitution, T866A substitution, and A877T substitution. IVT of different RTE drivers and reporters in cis configuration was carried out as described above. U20S cells were prepared as described above.

[0770] These results illustrate several important aspects disclosed herein. First, replacement of the native PCNA recruitment motif in R4-AC with the equivalent domain from P21 significantly improved activity, possibly by improving recruitment of PCNA. Second, without wishing to be bound to theory, improving interaction of DNA binding domain through addition of positive charge (Q195K) or improving interaction of reverse transcriptase domain to RNA/DNA through addition of a hydrogen bond (A877T) may improve one or more of efficiency, specificity, accuracy, fidelity or processivity of the engineered RTE.

Example 4: Drivers comprising *SpuFz1* nuclease R2-TG fusions

[0771] Nucleic acids were designed and produced to encode non-limiting examples of engineered proteins which serve as drivers comprising *SpuFz1* nuclease R2-TG fusions and include constructs having amino acid sequences set forth in SEQ ID NOS: 160-172.

[0772] These were tested and evaluated using an experimental transgene integration assay as described in Example 2, with the following changes:

[0773] The trans driver construct contains a R2 polypeptide-Fanzor fusion and with or without a polyadenylated SV40 sequence ending in A30-N10-A70 (wherein the N10 is a non-polyA

sequence). RNA versions of the reporter and driver constructs produced via in vitro transcription (IVT) as described above.

[0774] Sample quantification and protocols for U2OS cells are described in Examples above. The engineered drivers, each including an N-terminal NLS are described, hereinbelow:

[0775] **SEQ ID NO: 160** sets forth the amino acid sequence of a SpuFz1 nuclease R2-TG fusion. The SpuFz1 portion is fused at the N-terminus via a linker to a C-terminal R2-TG protein.

[0776] **SEQ ID NO: 161** sets forth the amino acid sequence of a SpuFz1 nuclease R2-TG fusion. The SpuFz1 portion is fused at the N-terminus via a linker to a truncated C-terminal R2-TG protein, lacking first 209 amino acid residues.

[0777] **SEQ ID NO: 162** sets forth the amino acid sequence of a SpuFz1 nuclease R2-TG fusion. The SpuFz1 portion is fused at the N-terminus via a linker to a truncated C-terminal R2-TG protein, with native endonuclease domain deleted.

[0778] **SEQ ID NO: 163** sets forth the amino acid sequence of a SpuFz1 nuclease R2-TG fusion. The SpuFz1 portion is fused at the C-terminus via a linker to a truncated N-terminal R2-TG protein, with native endonuclease domain deleted.

[0779] **SEQ ID NO: 164** sets forth the amino acid sequence of a SpuFz1 nuclease R2-TG fusion. The SpuFz1 portion is fused at the C-terminus via a linker to an N-terminal R2-TG protein, carrying a K1307A mutation.

[0780] **SEQ ID NO: 165** sets forth the amino acid sequence of a dead SpuFz1 nuclease R2-TG fusion. The SpuFz1 carries a D606A mutation and is fused at the N-terminus fused via a linker to a C-terminal R2-TG protein.

[0781] **SEQ ID NO: 166** sets forth the amino acid sequence of a dead SpuFz1 nuclease R2-TG fusion. The SpuFz1 carries a D606A mutation and is fused at the N-terminus via a linker to a C-terminal R2-TG protein, lacking first 209 amino acid residues.

[0782] SEQ ID NO: 167 sets forth the amino acid sequence of a SpuFz1 nuclease R2-TG fusion. The SpuFz1 is fused at the N-terminus via a linker to a C-terminal R2-TG protein (lacking 209 N-terminal amino acids).

[0783] SEQ ID NO: 168 sets forth the amino acid sequence of a SpuFz1 nuclease R2-TG fusion. The SpuFz1 is fused at the N-terminus via a linker to a C-terminal R2-TG protein without the native DNA binding domain, and carrying K1307A mutation.

[0784] SEQ ID NO: 169 sets forth the amino acid sequence of a SpuFz1 nuclease R2-TG fusion. The SpuFz1 is fused at the N-terminus via a linker to a C-terminal R2-TG protein carrying a K1307A mutation.

[0785] SEQ ID NO: 170 sets forth the amino acid sequence of a SpuFz1 nuclease R2-TG fusion. The SpuFz1 is fused at the C-terminus via a linker to an N-terminal R2-TG protein (lacking 209 N-terminal amino acid residues).

SEQ ID NO: 171 sets forth the amino acid sequence of a SpuFz1 nuclease R2-TG fusion. The SpuFz1 is fused at the C-terminus via a linker to a truncated N-terminal R2-TG protein.

Example 5: Engineered proteins comprising R2-TG chimeras or modified RT domains

[0786] Nucleic acids were designed and produced to encode non-limiting examples of engineered proteins comprising an R2-TG chimeras or fusions with modified reverse transcriptase (RT) domains. These were tested and evaluated using an experimental transgene integration assay described in Example 2. Non-limiting examples of the heterologous polypeptides include the following modifications: R2-TG chimera with a RT domain or a RNA binding domain (RBD) set forth in any one of SEQ ID NOS 290-301 are as follows: R2-TG chimera with Vingi RT (SEQ ID NO: 270), R2-TG chimera with MMLV RT (SEQ ID NO: 271), R2-TG chimera with R2NS RT (SEQ ID NO: 272), R2-TG chimera with L2_24_ACar-RT (SEQ ID NO: 273), R2-TG chimera with CR1-10_Ami-RT (SEQ ID NO: 274), R2-TG chimera with R2Toc-RT (SEQ ID NO: 275), R2-TG chimera with ZFI2-2 RT (SEQ ID NO: 276), R2-TG chimera with Marathon RT (SEQ ID NO: 277), R2-TG chimera with TGIRT_III RT (SEQ ID NO: 278), R2-TG mutated RT with YADD RT motif (SEQ ID NO: 279), R2-TG with RT deletion (SEQ ID NO: 280), R2-TG internal fusion with MMLV RT derived peptide in the RT domain (SEQ ID NO: 281), R2-TG chimera with mutated R2-TG and MMLV RT

sequence from SEQ ID NO: 281 (SEQ ID NO: 282), R2-TG chimera with SSIV RT (SEQ ID NO: 283), mutated R2-TG fusion with SSIV RT (SEQ ID NO: 284), R2-TG internal fusion with N-terminal fused RNA binding domain (RBD) motif (SEQ ID NO: 285), R2-TG internal fusion of RBD motif (SEQ ID NO: 286).

[0787] These were tested and evaluated using an experimental transgene integration assay described in Example 2.

Example 6: Engineered transgenes comprising 3' terminal RNA pseudoknot

[0788] Gene delivery constructs (reporters) were designed and produced exemplifying engineered transgenes comprising 3' terminal RNA pseudoknot, replacing the polyadenylation signal and polyA tail. The GFP transgene may be replaced with a gene to be delivered and incorporated into a subject's cell, for example into the genome of a mammalian cell. Non-limiting examples of these RNA pseudoknot constructs for in trans administration include constructs having nucleic acid sequences set forth in SEQ ID NOS: 180-182:

[0789] SEQ ID NO: 180 is a DNA sequence of an R2-TG reporter with GFP transgene and EvopreQ1 pseudoknot replacing polyA tail. GFP transgene cassette (anti-sense, with CMV promoter and HSV TK polyadenylation sequence).

[0790] SEQ ID NO: 181 is a DNA sequence of an R2-TG reporter with GFP transgene and Mpknot1 pseudoknot replacing polyA tail. GFP transgene cassette (anti-sense, with CMV promoter and HSV TK polyadenylation sequence).

[0791] SEQ ID NO: 182 is a DNA sequence of an R2-TG reporter with GFP transgene and MALAT1 pseudoknot replacing polyA tail. GFP transgene cassette (anti-sense, with CMV promoter and HSV TK polyadenylation sequence).

[0792] The R2-TG wild-type driver (SEQ ID NO: 1) and GFP reporter in trans configuration were tested as described above, and compared with pseudoknot reporters. IVT of different RTE drivers and reporters in trans configuration was carried out as described in Examples above in U20S cells. Results shown here are from two technical repeats.

[0793] Results are shown in FIG. 8. The reporters comprised modified R2-TG transgene constructs as described above. The constructs are as follows, from left to right: cell control, R2-

TG polyadenylated reporter control, R2-TG EvopreQ1 reporter control, Mpkn1 reporter control, R2-TG MALAT1 reporter control, R2-TG polyadenylated reporter with driver, R2-TG EvopreQ1 reporter with driver, Mpkn1 reporter with driver, R2-TG MALAT1 reporter with driver.

[0794] Significant activity (>10% above control) was observed with all three pseudoknot constructs, and equivalent activity as the default polyadenylated reporter was observed with the MALAT1 pseudoknot. This indicates that the polyadenylated tail of the RNA reporter is dispensable and may be replaced with a shorter pseudoknot.

Example 7: The ORF1-like RNA binding protein (ORBP) system

[0795] Nucleic acids were designed and produced exemplifying engineered transgenes comprising an open reading frame for cis-acting RNA binding proteins. Non-limiting examples of these constructs include constructs having nucleic acid sequences set forth in SEQ ID NOS: 190-204. The following constructs comprise a retrotransposable element reporter and ORF1p element and contain a GFP transgene cassette. The GFP transgene cassette may be replaced with any gene to be delivered and incorporated into a subject cell, for example into the genome of a mammalian cell. In some embodiments, the constructs include the following elements, 5' to 3': Upstream (e.g., 5') homology arm- 5' RTE-UTR (e.g., from R2-TG)-an ORF1p element-optional tetraloop-GFP "transgene"-3' RTE-UTR (e.g., R2-TG)-downstream (e.g., 3' homology arm). The GFP transgene may be replaced with a transgene (i.e., gene of interest). Without wishing to be bound to theory, the optional tetraloop (SEQ ID NOS: 199-204) may function as a reverse transcriptase (RT) inhibitor.

[0796] Constructs generated were as follows:

[0797] **SEQ ID NO: 190** sets forth a nucleic acid sequence of an R2-TG reporter with GFP transgene and LINE1 ORF1p from *Mus musculus* with R2 5' and 3' UTRs and a CMV promoter and HSV TK polyadenylation sequence.

[0798] **SEQ ID NO: 191** sets forth a nucleic acid sequence of an R2-TG reporter with GFP transgene and ORF1p from ZFL2-1 RTE from zebrafish with R2 5' and 3' UTRs and a CMV promoter and HSV TK polyadenylation sequence.

[0799] SEQ ID NO: 192 sets forth a nucleic acid sequence of an R2-TG reporter with GFP transgene and ORF1p from DOC2 RTE from *Drosophila* with R2 5' and 3' UTRs and a CMV promoter and HSV TK polyadenylation sequence.

[0800] SEQ ID NO: 193 sets forth a nucleic acid sequence of an R2-TG reporter with GFP transgene and ORF1p from Maui RTE from Fugu fish with R2 5' and 3' UTRs and a CMV promoter and HSV TK polyadenylation sequence.

[0801] SEQ ID NO: 194 sets forth a nucleic acid sequence of an R2-TG reporter with GFP transgene and ORF1p from I3 RTE from Zebrafish with R2 5' and 3' UTRs and a CMV promoter and HSV TK polyadenylation sequence.

[0802] SEQ ID NO: 195 sets forth a nucleic acid sequence of an R2-TG reporter with GFP transgene and LINE1 ORF1p from *Mus musculus* with three consecutive stop codons (e.g., tgataatag).

[0803] SEQ ID NO: 196 sets forth a nucleic acid sequence of an R2-TG reporter with GFP transgene and ORF1p from ZFL2-1 RTE from zebrafish. The ORF1p includes three stop codons.

[0804] SEQ ID NO: 197 sets forth a nucleic acid sequence of an R2-TG reporter with GFP transgene and ORF1p from DOC2 RTE from *Drosophila* with three stop codons.

[0805] SEQ ID NO: 198 sets forth a nucleic acid sequence of an R2-TG reporter with GFP transgene and ORF1p from Maui RTE from Fugu fish with three stop codons.

[0806] SEQ ID NO: 199 sets forth a nucleic acid sequence of an R2-TG reporter with GFP transgene and ORF1p from I3 RTE from Zebrafish, ORF1p has three stop codons. A reverse transcriptase (RT) inhibiting sequence is present between ORF1p and the GFP transgene cassette.

[0807] SEQ ID NO: 200 sets forth a nucleic acid sequence of an R2-TG reporter with GFP transgene and LINE1 ORF1p from *Mus musculus*. An RT inhibiting sequence is present between ORF1p and the GFP transgene cassette.

[0808] SEQ ID NO: 201 sets forth a nucleic acid sequence of an R2-TG reporter with GFP transgene and ORF1p from ZFL2-1 RTE from zebrafish. An RT inhibiting sequence is present between the ORF1p and the GFP transgene cassette.

[0809] SEQ ID NO: 202 sets forth a nucleic acid sequence of an R2-TG reporter with GFP transgene and ORF1p from DOC2 RTE from Drosophila. An RT inhibiting sequence is present between ORF1p and the GFP transgene cassette

[0810] SEQ ID NO: 203 sets forth a nucleic acid sequence of an R2-TG reporter with GFP transgene and ORF1p from Maui RTE from Fugu fish. An RT inhibiting sequence is present between ORF1p and the GFP transgene cassette

[0811] SEQ ID NO:204 sets forth a nucleic acid sequence of an R2-TG reporter with GFP transgene and ORF1p (lowercase text) from I3 RTE from Zebrafish. An RT inhibiting sequence is present between ORF1p and the GFP transgene cassette.

[0812] These modified reporters with ORF1p in cis improve integration efficiency when tested in the assay systems described above. Without wishing to be bound by theory, ORF1p protein may bind to the RNA reporter, facilitating RNA nuclear import and RNA stability.

Example 8: Transgene integration in human primary hepatocytes

[0813] RTE variants were tested in human primary hepatocytes using the following methods: Primary hepatocytes were isolated from two different donors and cryopreserved. Cryopreserved plateable primary human hepatocytes were grown on collagen-coated cell culture plates (Collagen type I solution, rat tail. Sigma). Thawing and plating media were purchased from Primacyt. Human Hepatocyte Basal Medium (Primacyt) was used after adding Supplements 1 and 2. The Enzyme Cell Detachment Medium for cell harvesting for FACS analysis (Accutase, ThermoFisher). Primary hepatocytes were transfected using lipid nanoparticles according to manufacturer's instructions (Ignite Nano Assembler, Precision Nanosystems).

[0814] LNP-mediated mRNA delivery: Cells were harvested for flow cytometry and dPCR analysis. The human hepatocytes were recovered and treated with mRNA formulated with LNP. At indicated time points, cells were harvested and taken for FACS analysis to assess integration level: Following incubation with LNP-mRNA formulations, the growth medium was aspirated

from the cell culture vessel and cells rinsed with PBS. Accutase solution was added to cover the cell monolayer (example: 100uL to the well of 24w-plate) and cells incubated at 37°C for 3-10 minutes. Enzyme activity was neutralized with x2 volume of warm PBS with 20% FBS. Half of the cell suspension was transferred to a 96w plate and flow cytometry analysis performed. IVT of different RTE drivers and reporters in trans configuration was carried out as described in Examples, above. Primary hepatocytes were treated with LNP-mRNA as described. Results shown here are from two repeats in hepatocytes. The remainder was kept for dPCR analysis using PCR strips.

[0815] FIGS. 9A and 9B show results of integration assays using drivers comprising peptide fusions and mutations in R2-TG protein (mutations with respect to SEQ ID NO: 1) in primary hepatocytes from Donor 1 and Donor 2, respectively. In this experiment, different retrotransposable element driver constructs were used with two different reporter constructs (GFP). The drivers were modified R2-TG protein constructs described above.

[0816] In FIG. 9A the constructs are as follows, from left to right: untreated cells, R2-TG reporter only control, R2-TG GFP Sirloin reporter control, R2-TG driver with Sirloin reporter, R2-TG driver with HAG 3'UTR and GFP reporter, R2-TG driver with 3 stop codons and GFP reporter, R2-TG driver with T1011S amino acid substitution and GFP reporter, R2-TG driver with W345R amino acid substitution and GFP reporter, R2-TG driver with 3' terminal PCNA motif and GFP reporter, R2-TG driver with D555K amino acid substitution and GFP reporter, codon optimized R2-TG driver (SEQ ID NO: 12) and GFP reporter, R2-TG control with GFP reporter, R2-TG driver with T1011S + I1219N+D555K+P280K amino acid substitutions and GFP reporter.

[0817] In FIG. 9B the constructs are as follows, from left to right: untreated cells, R2-TG reporter only control, R2-TG GFP Sirloin reporter control, R2-TG driver with Sirloin reporter, R2-TG driver with HAG 3'UTR and GFP reporter, R2-TG driver with 3 stop codons and GFP reporter, R2-TG driver with T1011S amino acid substitution and GFP reporter, R2-TG driver with W345R amino acid substitution and GFP reporter, R2-TG driver with 3' terminal PCNA motif and GFP reporter, R2-TG driver with D555K amino acid substitution and GFP reporter, codon optimized R2-TG driver (SEQ ID NO: 11) and GFP reporter, R2-TG control, R2-TG driver with T1011S + I1219N+D555K+P280K amino acid substitutions and GFP reporter.

[0818] Moderate improvements to transgene integration by LNP delivery of mRNA primary human hepatocytes were observed by stabilizing the R2-TG driver mRNA with HAG 3' UTR, by adding a sirloin sequence to the R2-TG GFP reporter mRNA, and with the T1011S mutation. We observe significant improvement with R2-TG driver with mutation D555K (~3-fold improvement across two donors), as well as with codon optimization (SEQ ID: 12). A combination variant of T1011S+I1219N+D555K+P280K improved transgene integration activity of R2-TG driver from 2 to 4-fold.

[0819] The substitution of aspartate to lysine at position 555 in R2-TG (reference SEQ ID NO: 1) improved transgene integration activity. An improvement of about 10%, 20%, 30%, 40% or at least 50% over control was observed. Accordingly, mutations (i.e., amino acid substitutions) to positively charged residues (e.g., arginine, histidine), hydrogen bond forming residues (e.g., serine, threonine, cysteine, glutamine, asparagine, tyrosine), aromatic residues (e.g., tyrosine, tryptophan, phenylalanine), or residues including alanine, leucine, valine, methionine, isoleucine), improve activity at this site.

Example 9: Novel R2 RTEs

[0820] Novel R2 RTEs encoding R2 proteins having amino acid sequences set forth in SEQ ID NOs: 401-480, as well as comprising RTE-UTR nucleotide sequences set forth in SEQ ID NOS: 500-712, were identified in a screen by scanning, utilizing a consensus sequence, for insertions of retrotransposable elements in the rDNA genes within genomes present in public databases.

Example 10: R2-1_TG engineered driver with reporters from different species

[0821] To potentially increase R2 integration, the R2-TG driver construct encoding the engineered R2-TG driver (SEQ ID NO: 636) was tested with R2 gene expression constructs comprising a MNDopt-driven GFP transgene with RTE-UTRs from R2 retroelements from several different species were tested. These experiments were run with TransAct activation to see a high signal. Donor 12 T cells were activated with TransAct, then transfected with the indicated mRNA-LNPs for 2 days, and FACS and digital PCR with MNDopt or GFP probes was performed 5 days post LNP transfection. Species tested were *Taeniopygia guttata* (TG), *Phylloscopus trochilus trochilus* (PTT), *Limnodromus scolopaceus* (LS), and *Melierax canorus* (MC).

[0822] As shown in FIG. 10A, the experimental conditions shown in the X-axis are to be interpreted as follows: NT = Non-treated, R only = gene expression construct (“reporter”) only negative control, D+R = driver construct + reporter. Error bars show SE of duplicates. Also as shown in FIG. 10A, TG = *Taeniopygia guttata*, PTT = *Phylloscopus trochilus trochilus*, LS = *Limnodromus scolopaceus*, MC = *Melierax canorus*. As such, for example, “R2_TG D + TG R” means that the cells were transfected with a EX5802 R2-TG driver construct and an R2 gene expression construct comprising RTE-UTRs from the same species that the driver polypeptide was derived from, namely R2-TG from *Taeniopygia guttata*. In another example, “R2_TG D + PTT R” means that the cells were transfected with a EX5802 R2-1 TG driver construct and an R2 gene expression construct comprising RTE-UTRs from a different species, namely R2-1_PTT from *Phylloscopus trochilus trochilus*. It was found, as shown in FIG. 10A, that the R2-1_PTT reporter was better than R2-1_TG reporter when used with the engineered R2-TG driver (SEQ ID NO: 636), while R2-1_LS and R2-1_MC were worse than R2-1_TG (FIG. 10A). The R2-1_PTT reporter comprises a 5' terminal region (SEQ ID NO: 20756) comprising a 5' RTE-UTR (SEQ ID NO: 20506) from the R2 retroelement found in *Phylloscopus trochilus trochilus* (R2-1_PTT), and a 3' terminal region (Seq ID NO: 20757), comprising a 3' RTE-UTR (SEQ ID NO: 20614) from R2-PTT and a PolyA signal.

[0823] As shown in FIG. 10B, the combination of the EX5802 R2-1 TG driver construct and the R2-1_PTT reporter were tested at several driver construct : reporter construct ratios (by weight of encapsulated RNA, with a constant total of 60ug per LNP), with a driver construct : reporter construct ratio of 1:1 being the control condition.

[0824] It was again found that the R2-1_PTT reporter was more effective in driving the GFP transgene expression than the R2-1_TG reporter. Expression levels were mostly the same across the different ratios. That said, the 2:1 driver:reporter construct ratio was somewhat higher, achieving 43 integrations of the MND probe and 25 integrations of the GFP probe by dPCR and 0.41% GFP+ by FACS.

Examples 11-78

[0825] In the following examples, RNA versions of driver constructs and reporter constructs were produced via IVT, transfections, and FACS analysis of GFP+ cells, were performed as generally described in Example 2. Further details regarding the Examples 11-78 are provided

below: Transfection was carried out with driver construct alone (negative control), reporter construct alone (negative control), with the driver and reporter constructs being delivered (control or test) with LipofectamineTM to HUH7 (human hepatic cell line) cells. Full-length integration was assessed by FACS at 24 hours. Descriptions of the experimental driver and reporter constructs are provided in each example.

[0826] The amino acid and nucleotide sequences referred to in these examples correspond to each other as follows: The SEQ ID NOS for the DNA sequences are systematically assigned to be greater by 10,000 compared to the SEQ ID NO for the amino acid sequence of the engineered driver that it encodes. As such, for example, the engineered driver having the sequence set forth in SEQ ID NO: 120 is encoded by a DNA having the sequence set forth in SEQ ID NO: 10120 (or by a corresponding RNA version thereof). In another example, the engineered driver having the sequence set forth in SEQ ID NO: 1976 is encoded by a DNA having the sequence set forth in SEQ ID NO: 11976 (or by a corresponding RNA version thereof).

[0827] The datasets described in the following Examples 11-78 hereinbelow included the following standard controls, unless otherwise stated:

[0828] Negative controls were LipofectamineTM-only and at least one of WT R2-TG reporter-only (SEQ ID NO:175) or a GFP reporter-only (with modified, ancestral RTE-UTRs; SEQ ID NO:581).

[0829] Three positive controls were used, as follows:

[0830] 1) a WT R2-TG system ("SEQ ID NO: 1 + SEQ ID NO: 175) with a driver construct (SEQ ID NO: 10001) encoding wildtype R2-TG driver (SEQ ID NO:1) and the WT GFP reporter (SEQ ID NO:175);

[0831] 2) an engineered R2-TG system ("SEQ ID NO: 636 + SEQ ID NO: 175") with a driver construct (SEQ ID NO:10636) encoding a reference engineered R2-TG driver and WT GFP reporter; and

[0832] 3) another engineered R2-TG system ("SEQ ID NO: 636 + SEQ ID NO: 581") with a driver construct encoding the reference engineered R2-TG driver (SEQ ID NO: 10636) and a GFP reporter (SEQ ID NO:581) with modified, ancestral RTE-UTRs.

[0833] Error bars indicate in the respective figures represent standard error of two replicates.

[0834] SEQ ID NO: 1 (encoded by SEQ ID NO: 10001) sets forth the amino acid sequence of wildtype (WT) R2-TG RTE, and serves as a reference driver.

[0835] SEQ ID NO: 175 sets forth the nucleic acid sequence of a R2 reporter comprising a GFP transgene cassette (anti-sense, with CMV promoter and HSV TK polyadenylation sequence. 3' R2-UTR SEQ ID NO: 1866), 5' R2-UTR (SEQ ID NO: 2886), and 20bp homology arms to a 28S site in human genome (upstream SEQ ID NO: 3011 and (downstream SEQ ID NO: 3012).

[0836] SEQ ID NO: 581 sets forth the nucleic acid sequence of a R2 reporter comprising a GFP transgene cassette (anti-sense, with CMV promoter and HSV TK polyA) with a modified 5' R2-UTR that was reconstructed from monophyletic group 2 node 5 whose leaves are R2-LS set forth in SEQ ID NO:1348, R2-XC set forth in SEQ ID NO:1351, R2-SM set forth in SEQ ID NO:1349 and R2-GP set forth in SEQ ID NO:1350.

[0837] SEQ ID NO: 636 (encoded by SEQ ID NO: 10636) sets forth the amino acid sequence of a R2-TG driver having an N-terminal truncation (up to residue V161) with N-terminal fusion of SV40 NLS, CTIP polypeptide, and HMGN1 polypeptides and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, and A1282G.

[0838] Where utilized, 'HAG' refers to mouse Hemoglobin alpha, '3'UTR HAG' refers to the mouse Hemoglobin Alpha 3'UTR (SEQ ID NO: 3025), and '5'UTR HAG' mouse Hemoglobin Alpha 5'UTR (SEQ ID NO:3026).

[0839] Unless otherwise indicated, where N-terminal fusions and C-terminal fusions are listed with multiple heterologous polypeptides, the heterologous polypeptides are listed in a N-terminal to C-terminal order. As such, for example in SEQ ID NO: 636 above, the "R2-TG driver with N-terminal fusion of SV40 NLS, CTIP polypeptide, and HMGN1 polypeptides" means that driver comprises, in an N-terminal to C-terminal order, the SV40 NLS, the CTIP polypeptide, and the HMGN1 polypeptide.

Example 11: Modified RTE-UTR sequences created through ancestral reconstruction of multiple R2 RTEs improved R2 transgene integration

[0840] Modified R2 RTE-UTR elements were created through ancestral reconstruction using an ancestral genome reconstruction program Ancestors 1.1, based on RTE-UTR elements from a plurality of R2 RTEs. The modified RTE-UTRs, which included modified 5' RTE-UTRs and/or 3' RTE-UTRs were based on a plurality of R2 RTEs. Different sets of R2 RTEs, as described herein below, were used to generate different modified RTE-UTRs based on ancestral reconstruction (see, e.g., Campitelli et al., Reconstruction of full-length LINE-1 progenitors from ancestral genomes, Genetics, Volume 221, Issue 3, July 2022, and Diallo AB et al., Ancestors 1.0: a web server for ancestral sequence reconstruction. Bioinformatics. 2010;26(1):130–131, whose ancestral reconstruction procedures are incorporated herewith).

[0841] The following reporters (i.e. gene delivery constructs) with modified 5' RTE-UTRs and/or 3' RTE-UTRs, created as described above, and having nucleic acid sequences set forth in any one of SEQ ID NO: 589-592 or 594-598 described hereinbelow, were generated:

[0842] SEQ ID NO: 589 sets forth the nucleic acid sequence of a GFP reporter comprising a transgene cassette (anti-sense, with CMV promoter and HSV TK polyA) with a modified R2 3' UTR (SEQ ID NO: 3003) reconstructed from monophyletic group 1 node 4 whose leaves are R2-MC set forth in SEQ ID NO:1338, R2-HN set forth in SEQ ID NO:1339 and R2-GA set forth in SEQ ID NO: 1340.

[0843] SEQ ID NO: 590 sets forth the nucleic acid sequence of a GFP reporter comprising a transgene cassette (anti-sense, with CMV promoter and HSV TK polyA) with a modified R2 3' UTR (SEQ ID NO: 3004) that was reconstructed from monophyletic group 2 node 5 whose leaves are R2-LS set forth in SEQ ID NO:1344, R2-SM set forth in SEQ ID NO:1345, R2-GP set forth in SEQ ID NO: 1346 and R2-XC set forth in SEQ ID NO:1347.

[0844] SEQ ID NO: 591 sets forth the nucleic acid sequence of a GFP reporter comprising a transgene cassette (anti-sense, with CMV promoter and HSV TK polyA) with a modified R2 3' UTR (SEQ ID NO: 3005) that was reconstructed from monophyletic group 3 node 7 whose leaves are R2-AC SEQ ID NO:1352, R2-CC SEQ ID NO:1353 and R2-CC SEQ ID NO:1354.

[0845] SEQ ID NO: 592 sets forth the nucleic acid sequence of a GFP reporter comprising a transgene cassette (anti-sense, with CMV promoter and HSV TK polyA) with modified R2 5' UTR (SEQ ID NO: 3006) that was reconstructed from monophyletic group 1 node 4 whose leaves are R2-MC SEQ ID NO:1441, R2-GA SEQ ID NO:1442 and R2-HN SEQ ID NO:1443.

[0846] SEQ ID NO: 581 sets forth the nucleic acid sequence of a GFP reporter comprising a transgene cassette (anti-sense, with CMV promoter and HSV TK polyA) with a modified R2 5' UTR (SEQ ID NO: 1874) that was reconstructed from monophyletic group 2 node 5 whose leaves are R2-LS set forth in SEQ ID NO:1348, R2-XC set forth in SEQ ID NO:1351, R2-SM set forth in SEQ ID NO:1349 and R2-GP set forth in SEQ ID NO:1350.

[0847] SEQ ID NO: 594 sets forth the nucleic acid sequence of a GFP reporter comprising a transgene cassette (anti-sense, with CMV promoter and HSV TK polyA) with a modified R2 5' UTR (SEQ ID NO: 3006) that was reconstructed from monophyletic group 3 node 4 whose leaves are R2-OM set forth in SEQ ID NO:1355, R2-ASS set forth in SEQ ID NO:1356 and R2-PTT set forth in SEQ ID NO:1357.

[0848] SEQ ID NO: 595 sets forth the nucleic acid sequence of a GFP reporter comprising a transgene cassette (anti-sense, with CMV promoter and HSV TK polyA) with a modified R2 5' UTR (SEQ ID NO: 3007) that was reconstructed from monophyletic group 4 node 6 whose leaves are R2-AT set forth in SEQ ID NO:1358 and R2-TG set forth in SEQ ID NO:1359.

[0849] SEQ ID NO: 596 sets forth the nucleic acid sequence of a GFP reporter comprising a transgene cassette (anti-sense, with CMV promoter and HSV TK polyA) with a modified R2 5' UTR (SEQ ID NO: 1846) that was reconstructed from monophyletic group 7 node 5 whose leaves are R2-SC set forth in SEQ ID NO:1363, R2-ZA set forth in SEQ ID NO:1364, R2-AC/R2-AN set forth in SEQ ID NO:1365.

[0850] SEQ ID NO: 597 sets forth the nucleic acid sequence of a GFP reporter comprising a transgene cassette (anti-sense, with CMV promoter and HSV TK polyA) with a modified R2 5' UTR (SEQ ID NO: 1847) that was reconstructed from monophyletic group 8m node 10 whose leaves are R2-ASS set forth in SEQ ID NO:1356 and R2-PTT set forth in SEQ ID NO:1357. The sequences were modified by trimming a long polyA sequence in the middle of the UTRs.

[0851] SEQ ID NO: 598 sets forth the nucleic acid sequence of a GFP reporter comprising a transgene cassette (anti-sense, with CMV promoter and HSV TK polyA) with a modified R2 5' UTR (SEQ ID NO: 1848) that was reconstructed from monophyletic group 8f node 10 whose leaves are R2-ASS set forth in SEQ ID NO:1356 and R2-PTT set forth in SEQ ID NO:1366.

[0852] The above reporter constructs were then tested for improvement in transgene integration. Each of the above-noted reporter constructs was delivered alone or in the presence of a driver construct encoding a reference engineered R2-TG driver polypeptide having an amino acid sequence set forth in SEQ ID NO: 636 and encoded by a nucleic acid sequence set forth in SEQ ID NO: 10636.

[0853] The dataset included the following controls: Lipofectamine-only control; GFP reporter having a nucleotide sequence set forth in SEQ ID NO:589-592 and 594-598 as negative controls, an engineered R2-TG system ("SEQ ID NO: 636 + SEQ ID NO: 175") as positive control. Error bars indicate standard error of two replicates.

[0854] FIGS. 11A and 11B are bar graphs showing results of an integration assay in USO2 cells transfected with a driver construct encoding the driver polypeptide having an amino acid sequence set forth in SEQ ID NO:636 and one of several different reporters (i.e. gene delivery construct) set forth in SEQ ID NOs: 589-592 or 594-598.

[0855] Four of the modified reporters (gene delivery constructs) exhibited activity above the reference reporter (SEQ ID NO:175): SEQ ID NO:592, SEQ ID NO:581, SEQ ID NO:597 and SEQ ID NO:598.

Example 12: Modified RTE-UTR sequences

[0856] The following reporters (i.e. gene delivery constructs) comprising additional modified 5' RTE-UTRs and/or 3' RTE-UTRs, created as described above and having nucleic acid sequences set forth in any one of SEQ ID NO: 599-614 described, hereinbelow, were generated and tested:

[0857] SEQ ID NO: 599 sets forth the nucleic acid sequence of a GFP reporter comprising a transgene cassette (anti-sense, with CMV promoter and HSV TK polyA) with a modified 5' UTR from R2-ATS (SEQ ID NO:20574)

[0858] SEQ ID NO: 600 sets forth the nucleic acid sequence of a GFP reporter comprising a transgene cassette (anti-sense, with CMV promoter and HSV TK polyA) with a modified 5' UTR from R2-EM (SEQ ID NO: 20597).

[0859] SEQ ID NO: 601 sets forth the nucleic acid sequence of a GFP reporter comprising a transgene cassette (anti-sense, with CMV promoter and HSV TK polyA) with a modified 5' UTR from R2-SP (SEQ ID NO:20586 and 1849).

[0860] SEQ ID NO: 602 sets forth the nucleic acid sequence of a GFP reporter comprising a transgene cassette (anti-sense, with CMV promoter and HSV TK polyA) with a modified 5' UTR from R2-SK (SEQ ID NO:20547 and 1850).

[0861] SEQ ID NO: 603 sets forth the nucleic acid sequence of a GFP reporter comprising a transgene cassette (anti-sense, with CMV promoter and HSV TK polyA) with a modified 5' UTR (SEQ ID NO:2883) that was reconstructed from monophyletic group 8m node 11 whose leaves are R2-OM SEQ ID NO:1355, R2-AT SEQ ID NO:1358 and R2-TG SEQ ID NO:1359. The sequences were modified by trimming a long PolyA sequence in the middle of the UTRs.

[0862] SEQ ID NO: 604 sets forth the nucleic acid sequence of a GFP reporter comprising a transgene cassette (anti-sense, with CMV promoter and HSV TK polyA) with modified 5' UTR from R2-ZYZ (SEQ ID NO: 2884)

[0863] SEQ ID NO: 605 sets forth the nucleic acid sequence of a GFP reporter comprising a transgene cassette (anti-sense, with CMV promoter and HSV TK polyA) with modified 5' UTR from R2-NN (SEQ ID NO: 20568).

[0864] SEQ ID NO: 606 sets forth sets forth the nucleic acid sequence of a GFP reporter comprising a transgene cassette (anti-sense, with CMV promoter and HSV TK polyA) with modified 5' UTR from R2-AO (SEQ ID NO:20569).

[0865] SEQ ID NO: 607 sets forth sets forth the nucleic acid sequence of a GFP reporter comprising a transgene cassette (anti-sense, with CMV promoter and HSV TK polyA) with modified 5' UTR from RTE R2- BL (SEQ ID NO:20558).

[0866] SEQ ID NO: 608 sets forth the nucleic acid sequence of a GFP reporter comprising a transgene cassette (anti-sense, with CMV promoter and HSV TK polyA) with modified 5' UTR from R2-AB (SEQ ID NO:20571).

[0867] SEQ ID NO: 609 sets forth the nucleic acid sequence of a GFP reporter comprising a transgene cassette (anti-sense, with CMV promoter and HSV TK polyA) with modified 5' UTR from RTE R2- AO (SEQ ID NO:20557 and 1871).

[0868] SEQ ID NO: 610 sets forth the nucleic acid sequence of a GFP reporter comprising a transgene cassette (anti-sense, with CMV promoter and HSV TK polyA) with modified 5' UTR from RTE R2-MA (SEQ ID NO:2885).

[0869] SEQ ID NO: 611 sets forth the nucleic acid sequence of a GFP reporter comprising a transgene cassette (anti-sense, with CMV promoter and HSV TK polyA) with modified 5' UTR from RTE R2- RP (SEQ ID NO:20567 and 1872).

[0870] SEQ ID NO: 612 sets forth the nucleic acid sequence of a GFP reporter comprising a transgene cassette (anti-sense, with CMV promoter and HSV TK polyA) with 6 repeats of the native R2-TG 5' RTE-UTR (SEQ ID NO:2886).

[0871] SEQ ID NO: 613 sets forth the nucleic acid sequence of a GFP reporter comprising a transgene cassette (anti-sense, with CMV promoter and HSV TK polyA) with the native R2-TG 5' RTE-UTR (SEQ ID NO:2886) and 3 repeats of the native R2-TG 3' UTR (SEQ ID NO:1866).

[0872] SEQ ID NO: 614 sets forth the nucleic acid sequence of a GFP reporter comprising a transgene cassette (anti-sense, with CMV promoter and HSV TK polyA) with 3 repeats of the native R2-TG 3' RTE-UTR (SEQ ID NO:2886).

[0873] FIGS. 12A and 12B are bar graphs showing the results of integration assays using the reference engineered R2-TG driver (SEQ ID NO: 636) and different reporters having nucleotide sequences set forth in SEQ ID NOs: 599-614. The dataset included the following controls: Lipofectamine-only control; GFP reporters (SEQ ID NO:599-614) as negative controls, R2-TG systems SEQ ID NO: 1 + SEQ ID NO: 175, SEQ ID NO: 636 + SEQ ID NO: 175, and SEQ ID NO: 636 + SEQ ID NO: 581 as positive controls.

[0874] Most of the modified reporters yielded results similar to the WT R2-TG control (SEQ ID NO: 1 + SEQ ID NO: 175). However, three reporters showed ~30% improvement compared to the engineered R2-TG system control (SEQ ID NO:636+SEQ ID NO:175): SEQ ID NO:601, SEQ ID NO: 602, and SEQ ID NO:608. Given that the same engineered R2-TG driver (SEQ ID NO: 636) was used in these experiments, the improvement is seen to be due to the modified 5' RTE-UTR sequence in each one of the reporters. One of these modified reporter (SEQ ID NO:602) exhibited a minor improvement over the engineered R2-TG system control (SEQ ID NO: 636 +SEQ ID NO:581).

Example 13

[0875] The following reporters (i.e. gene delivery constructs) having nucleic acid sequences set forth in any one of SEQ ID NO: 609 and 611-617 described, hereinbelow, were generated to test if having multiple copies of a RTE-UTR in a gene delivery construct would improve integration of the transgene comprised in the construct:

[0876] SEQ ID NO: 609 with a modified 5' UTR from RTE R2-AO is described hereinabove.

[0877] SEQ ID NO: 611 with a modified 5' UTR from RTE R2-MA is described hereinabove.

[0878] SEQ ID NO: 612 with 6 repeats of the native R2-TG 5' RTE-UTR is described hereinabove.

[0879] SEQ ID NO:613 with 3 repeats of the native R2-TG 5' RTE-UTR is described hereinabove.

[0880] SEQ ID NO: 615 sets forth GFP reporter comprising a transgene cassette (anti-sense, with CMV promoter and HSV TK polyA) with 2 repeats of the native R2-TG 3' RTE-UTR SEQ ID NO: 1866.

[0881] SEQ ID NO: 616 sets forth GFP reporter comprising a transgene cassette (anti-sense, with CMV promoter and HSV TK polyA) with 3 repeats of the native R2-TG 3' RTE-UTR SEQ ID NO: 1866.

[0882] SEQ ID NO: 614 sets forth GFP reporter comprising a transgene cassette (anti-sense, with CMV promoter and HSV TK polyA) with 3 repeats of the native R2-TG 5' RTE-UTR SEQ ID NO: 2886.

[0883] SEQ ID NO: 617 sets forth GFP reporter comprising a transgene cassette (anti-sense, with CMV promoter and HSV TK polyA) with 5'UTR HAG (SEQ ID NO: 3026).

[0884] FIGS. 13A and 13B is a bar graph showing the results of integration assays using the reference engineered R2-TG driver polypeptide (SEQ ID NO: 636) and different reporters having nucleotide sequences set forth in SEQ ID NOs: 609 and 611-617.

[0885] Controls were as described above.

[0886] All the modified reporters (+ SEQ ID NO 636) exhibited higher activity than the engineered R2-TG system with driver construct encoding reference engineered SEQ ID NO: 636 + SEQ ID NO: 175. That said, the other reference R2-TG system with driver construct encoding reference engineered R2-TG driver + GFP reporter with modified, ancestral RTE-UTRs (SEQ ID NO: 636 + SEQ ID NO: 581) exhibited the highest activity out of the cohort shown in FIGS. 13A and 13B.

Example 14: Combinations of modifications and fusions in engineered drivers enhance integration of the transgene.

[0887] Driver constructs encoding engineered R2-TG drivers having an amino acid sequence set forth in SEQ ID NOS: 636, 671, 675, 715, 838, 845, 906, 909, 1233-1236, 1240, 1249, 1251-1254 described hereinbelow were generated and tested with the GFP reporter construct having a nucleic acid sequence set forth in SEQ ID NO: 581.

[0888] SEQ ID NO:675 (encoded by SEQ ID NO: 10675) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue E184) with N-terminal fusion of a cMyc NLS, UL42 polypeptide, CTIP polypeptide, HMGN1 polypeptide; and C-terminal fusion of CHV core peptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G.

[0889] SEQ ID NO: 671 (encoded by SEQ ID NO: 11253) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue E218) with N-terminal

fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, E399M, D555K, D923K, T1011S, I1219N, A1282G. The DNA encoding the polypeptide is codon optimized for high frequency codons.

[0890] SEQ ID NO:1236 (encoded by SEQ ID NO: 11236) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue E184) with N-terminal fusion of a SV40 NLS, Sto7D polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G. The DNA encoding the polypeptide is codon-optimized for fast-translated codons.

[0891] SEQ ID NO:1235 (encoded by SEQ ID NO: 11235) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue E184) with N-terminal fusion of a cMyc NLS, Sto7D polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G. The DNA encoding the polypeptide is codon-optimized for fast-translated codons.

[0892] SEQ ID NO:1234 (encoded by SEQ ID NO: 11234) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue E184) with N-terminal fusion of a SV40 NLS, Tudor polypeptide, Sto7D polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G. The DNA encoding the polypeptide is codon-optimized for fast-translated codons.

[0893] SEQ ID NO:906 (encoded by SEQ ID NO: 10906) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue E184) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, F782Y, D923K, T1011S, I1219N, and A1282G. The DNA encoding the polypeptide is codon-optimized for fast-translated codons.

[0894] SEQ ID NO:909 (encoded by SEQ ID NO: 0909) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue E184) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, Y811L, D923K, T1011S, I1219N, and A1282G. The DNA encoding the polypeptide is codon-optimized for fast-translated codons.

[0895] SEQ ID NO:838 (encoded by SEQ ID NO: 10838) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue E184) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, V835I, D923K, T1011S, I1219N, and A1282G. The DNA encoding the polypeptide is codon-optimized for fast-translated codons.

[0896] SEQ ID NO:845 (encoded by SEQ ID NO: 10845) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue E184) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, P843R, D923K, T1011S, I1219N, and A1282G. The DNA encoding the polypeptide is codon-optimized for fast-translated codons.

[0897] SEQ ID NO:1254 (encoded by SEQ ID NO: 11254) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions E214R, N269R, P280K, E399M, D537Y, D555K, K700D, S895K, D923K, T989L, T1011S, L1013R, I1219N and A1282G. The DNA encoding the polypeptide is codon optimized for high frequency codons.

[0898] SEQ ID NO:1249 (encoded by SEQ ID NO: 11249) sets forth the amino acid sequence of a R2-TG polypeptide with amino acid substitutions Y1311A and R1337A. The DNA encoding the polypeptide is codon-optimized for fast-translated codons.

[0899] SEQ ID NO:1233 (encoded by SEQ ID NO: 11233) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, I683L, T1011S, I1219N, and A1282G.

[0900] SEQ ID NO:1252 (encoded by SEQ ID NO: 11252) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions G257R, P280K, N363H, E399M, A442P, V539G, K547R, D555K, D923K, T989L, T1011S, R1177G, K1085R, I1219N and A1282G. The DNA encoding the polypeptide is codon optimized for high frequency codons.

[0901] SEQ ID NO:1240 (encoded by SEQ ID NO: 11240) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue E184) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, F279Y, D923K, T1011S, I1219N, and A1282G.

[0902] SEQ ID NO:1251 (encoded by SEQ ID NO: 11251) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions G257R, P280K, N363H, E399M, H423Q, A442P, V539G, K547R, D555K, D923K, T989L, T1011S, R1177G, K1085R, I1219N, and A1282G.

[0903] SEQ ID NO:1253 (encoded by SEQ ID NO:11253) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions G257R, P280K, N363H, E399M, A442P, V539G, K547R, D555K, D923K, T989L, T1011S, K1085R, I1219N, A1282G.

[0904] SEQ ID NO:715 (encoded by SEQ ID NO: 10715) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, and D923K. The nucleic acid sequence encoding the polypeptide is set forth in SEQ ID NO: 10715.

[0905] SEQ ID NO: 636 (encoded by SEQ ID NO: 10636) is described supra.

[0906] The dataset included the following controls: Lipofectamine-only and GFP reporters (SEQ ID NO:175 and SEQ ID NO:581 as negative controls; and WT R2-TG control system SEQ ID NO: 1 + SEQ ID NO: 175 as positive control. Error bars indicate standard error of two replicates.

[0907] FIGS. 14A and 14B are bar graphs showing results of integration assays using engineered R2-TG drivers comprising heterologous polypeptide fusions, as well as amino acid deletions (N-terminal truncation up to amino acid V161 or E184) and amino acid substitutions in the R2-TG element, with each of the drivers having one of the amino acid sequences set forth in SEQ ID NOs: 636, 671, 675, 715, 838, 845, 906, 909, 1233-1236, 1240, 1249 and 1251-1254.

[0908] Most of the engineered drivers exhibited improved activity above that of WT R2-TG control, for example the drivers having an amino acid sequence set forth in any one of SEQ ID NOS: 636, 671, 675, 715, 1233-1236, 1240, 1251-1254.

[0909] In particular, the driver having the amino acid sequence set forth in SEQ ID NO:1234 (encoded by SEQ ID NO: 11234), which included the RNA/DNA processing polypeptide TUDOR (recruits RNA helicase DHX9), as well as processivity factor Sto7D, chromatin binding domain HMGN1, SV40 NLS, and amino acid substitutions P280K, D555K, T1011S, I1219N, and A1282G exhibited ~20% higher full-length gene integration than SEQ ID NO:1236, which does not have the TUDOR domain peptide.

[0910] SEQ ID NOS: 1234, 1235 and 1236 all include a Sto7D peptide N-terminal fusion, as well as N-terminal fusion of NLS (SV40 or cMyc) and HMGN1 peptides and amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G. Compared to SEQ ID NO: 636, with a similar set of N-terminal polypeptides aside from fusion of CTIP instead of Sto7D, these Sto7D-comprising drivers (SEQ ID NOS: 1234, 1235 and 1236) exhibited approximately 4-fold improved activity, indicating that N-terminal Sto7D fusion can unexpectedly improve activity .

Example 15: Modifications to the R2-TG RTE improved R2 transgene integration

[0911] Driver constructs encoding additional engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOS: 636, 678, 782, 783, 1244, 1245, 1257, 1258, 1262-1265, 1268-1271 described hereinbelow were generated and tested with the GFP reporter construct having a nucleic acid sequence set forth in SEQ ID NO: 581.

[0912] SEQ ID NO: 636 (encoded by SEQ ID NO: 10636) is described, supra.

[0913] SEQ ID NO: 1258 (encoded by SEQ ID NO: 11258) sets forth the amino acid sequence of a R2-TG RTE having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, K1157A, I1219N, and A1282G.

[0914] SEQ ID NO: 783 (encoded by SEQ ID NO: 10783) sets forth the amino acid sequence of a R2-TG RTE having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, EIF4A polypeptide, CTIP polypeptide, HMGN1 polypeptide and further includes

amino acid substitutions P280K, D555K, T1011S, I1219N, and A1282G. The driver of SEQ ID NO: 1258 differs from the driver of SEQ ID NO: 636 in the addition of the EIF4A polypeptide in the N-terminal fusion.

[0915] SEQ ID NO: 1271 (encoded by SEQ ID NO: 11271) sets forth the amino acid sequence of a R2-TG RTE having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, EIF4A polypeptide, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, and A1282G, and reverse transcriptase (RT) domain (AA residues 591-1029) replaced by corresponding residues from R2-PTT.

[0916] SEQ ID NO: 1270 (encoded by SEQ ID NO: 11270) sets forth the amino acid sequence of a R2-TG RTE having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, EIF4A polypeptide, CTIP polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, and A1282G.

[0917] SEQ ID NO: 782 (encoded by SEQ ID NO: 10782) sets forth the amino acid sequence of a R2-TG RTE having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, UL42 polypeptide, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, and A1282G.

[0918] SEQ ID NO:1269 (encoded by SEQ ID NO: 11269) sets forth the amino acid sequence of a R2-TG RTE having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, UL42 polypeptide, CTIP polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, and A1282G.

[0919] SEQ ID NO:1268 (encoded by SEQ ID NO: 11268) sets forth the amino acid sequence of a R2-TG RTE having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, UL42 polypeptide, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, and A1282G, and the reverse transcriptase domain (aa residues 591-1029) was replaced by the corresponding residues from R2-PTT.

[0920] SEQ ID NO:1265 (encoded by SEQ ID NO: 11265) sets forth the amino acid sequence of a R2-TG RTE having an N-terminal truncation (up to residue V161) with N-terminal fusion

of a SV40 NLS, UL42 polypeptide, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, Y571H, T1011S, I1219N, and A1282G.

[0921] SEQ ID NO: 1264 (encoded by SEQ ID NO: 11264) sets forth the amino acid sequence of a R2-TG RTE having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, UL42 polypeptide, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, and A1282G. The mRNA encoding this construct further includes a 5' UTR from human beta-globin (SEQ ID NO:206).

[0922] SEQ ID NO:1263 (encoded by SEQ ID NO: 11263) sets forth the amino acid sequence of a R2-TG RTE having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, UL42 polypeptide, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, and A1282G.

[0923] SEQ ID NO: 1262 (encoded by SEQ ID NO: 11262) sets forth the amino acid sequence of a R2-TG RTE having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, UL42 polypeptide, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, Y571H, D923K, T1011S, I1219N and A1282G.

[0924] SEQ ID NO:1836 (encoded by SEQ ID NO: 11836) sets forth the amino acid sequence of a R2-TG RTE having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, BRCA2 peptide, UL42 polypeptide, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, and A1282G.

[0925] SEQ ID NO:1257 (encoded by SEQ ID NO: 11257) sets forth the amino acid sequence of a R2-TG RTE having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, UL42 polypeptide, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, K1157A, I1219N, and A1282G.

[0926] SEQ ID NO:678 (encoded by SEQ ID NO: 10678) sets forth the amino acid sequence of a R2-TG RTE having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, TUDOR peptide, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, K1157A, I1219N, and A1282G.

[0927] SEQ ID NO:1245 (encoded by SEQ ID NO: 11245) sets forth the amino acid sequence of a R2-TG RTE having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, guinea pig cytomegalovirus UL42 polypeptide, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, K1157A, I1219N, and A1282G.

[0928] SEQ ID NO:1244 (encoded by SEQ ID NO: 11244) sets forth the amino acid sequence of a R2-TG RTE having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, IF4G 4A binding peptide, UL42 polypeptide, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, K1157A, I1219N, and A1282G.

[0929] The dataset included the following controls: Lipofectamine-only and GFP-only reporters (SEQ ID NO:175) and (SEQ ID NO:581) as negative controls; and wildtype R2-TG system (SEQ ID NO: 1 + SEQ ID NO: 175) as positive control. Error bars indicate standard error of two replicates.

[0930] FIGS. 15A and 15B is a bar graph showing results of integration assays using engineered R2-TG drivers comprising 5' truncations (to amino acid V161 or E184), fusion of heterologous proteins and amino acid substitutions in the R2-TG driver having an amino acid sequence set forth in SEQ ID NOs: 636, 678, 782, 783, 1244, 1245, 1257, 1258, 1262-1265, 1268-1271.

[0931] Most of the drivers exhibited activity above the control system. For example, SEQ ID NO:783, SEQ ID NO: 636, SEQ ID NO: 783, SEQ ID NO: 1271, SEQ ID NO: 1270, SEQ ID NO: 782, SEQ ID NO: 1269, SEQ ID NO: 1265, SEQ ID NO: 1264, SEQ ID NO: 1263, SEQ ID NO:1262, SEQ ID NO:, SEQ ID NO: 678, SEQ ID NO: 1245, and SEQ ID NO: 1244 exhibited improved activity above that of wild-type R2-TG (SEQ ID NO:1). In particular, SEQ ID NO: 783 (encoded by SEQ ID NO: 10783) which included the EIF4A polypeptide, as well as CTIP, HMGN1, SV40 NLS, and amino acid substitutions P280K, D555K, T1011S, I1219N, and A1282G and SEQ ID NO:678 (encoded by SEQ ID NO: 10678) which includes the TUDOR domain peptide, CTIP, HMGN1, SV40 NLS, and amino acid substitutions P280K, D555K, T1011S, I1219N, and A1282G. This indicates that fusing a cellular RNA helicase to the driver or recruiting a cellular RNA helicase to the RTE can improve integration activity. The driver set forth in SEQ ID NO: 1270 (encoded by SEQ ID NO: 11270) having EIF4A, CTIP, SV40 NLS,

and amino acid substitutions P280K, D555K, T1011S, I1219N, and A1282G was also more active than wild-type R2-TG driver, indicating the activity-improving effect of EIF4A.

Example 16

[0932] Driver constructs encoding engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOS: 636, 639, 648-650, 782, 1243, 1246, 1248, 1255, 1256, 1259, 1261, 1266, 1267, 1837 described, hereinbelow were generated and tested with the GFP reporter construct having a nucleic acid sequence set forth in SEQ ID NO: 581.

[0933] SEQ ID NO: 636 is described herein, supra.

[0934] SEQ ID NO:639 (encoded by SEQ ID NO: 10639) sets forth the amino acid sequence of a R2-TG RTE having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, and A1282G, and the R2-TG PCNA interaction motif replaced with LINE-1 PCNA interaction motif.

[0935] SEQ ID NO:648 (encoded by SEQ ID NO: 10648) sets forth the amino acid sequence of a R2-TG RTE having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, G658R, D923K, T1011S, I1219N, and A1282G.

[0936] SEQ ID NO:649 (encoded by SEQ ID NO: 10649) sets forth the amino acid sequence of a R2-TG RTE having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, G642R, D923K, T1011S, I1219N, and A1282G.

[0937] SEQ ID NO:650 (encoded by SEQ ID NO: 10650) sets forth the amino acid sequence of a R2-TG RTE having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, F973R, T1011S, I1219N, and A1282G.

[0938] SEQ ID NO:1243 (encoded by SEQ ID NO: 11243) sets forth the amino acid sequence of a R2-TG RTE having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, UL42 polypeptide, CTIP polypeptide, HMGN1 polypeptide, and C-terminal

fusion of EIF4A binding peptide derived from EIF4G and further includes amino acid substitutions P280K, D555K, D923K, T1011S, K1157A, I1219N, and A1282G.

[0939] SEQ ID NO:782 (encoded by SEQ ID NO: 10782) sets forth the amino acid sequence of a R2-TG RTE having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, UL42 polypeptide, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, and A1282G.

[0940] SEQ ID NO:1267 (encoded by SEQ ID NO: 11267) sets forth the amino acid sequence of a R2-TG RTE having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, UL42 polypeptide, CTIP polypeptide, and HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, and A1282G. The coding sequence has been codon optimized for fast-translated codons.

[0941] SEQ ID NO:1837 (encoded by SEQ ID NO: 11837) sets forth the amino acid sequence of a R2-TG RTE having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, UL42 polypeptide, CTIP polypeptide, and HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, and A1282G. The 3'UTR (human beta globin derived) has a nucleic acid sequence set forth in SEQ ID NO:207.

[0942] SEQ ID NO:1266 (encoded by SEQ ID NO: 11266) sets forth the amino acid sequence of a R2-TG RTE having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, UL42 polypeptide, CTIP polypeptide, and HMGN1 polypeptide and further includes amino acid substitutions G257R, N363H, E399M, V539G, H423Q, A442P, D537Y, K547R, D555K, R669L, N889G, F973E, T1011A, Q1019R, D1133P, K1085R, R1177G, Y1242K, and A1282G.

[0943] SEQ ID NO:1261 (encoded by SEQ ID NO: 11261) sets forth the amino acid sequence of a R2-TG RTE having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, UL42 polypeptide, CTIP polypeptide, and HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, Y571H, V660R, D923K, T1011S, I1219N, and A1282G.

[0944] SEQ ID NO:1259 (encoded by SEQ ID NO: 11259) sets forth the amino acid sequence of a R2-TG RTE having an N-terminal truncation (up to residue V161) with N-terminal fusion

of a SV40 NLS, TUDOR peptide, UL42 polypeptide, CTIP polypeptide, and HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, K1157A, I1219N, and A1282G.

[0945] SEQ ID NO:1256 (encoded by SEQ ID NO: 11256) sets forth the amino acid sequence of a R2-TG RTE having an N-terminal truncation (up to residue V161) with N-terminal fusion of a bipartite SV40 NLS, HCV core peptide, UL42 polypeptide, CTIP polypeptide, and HMGN1 polypeptide and further includes P280K, D555K, D923K, T1011S, I1219N, and A1282G.

[0946] SEQ ID NO:1255 (encoded by SEQ ID NO: 11255) sets forth the amino acid sequence of a R2-TG RTE having an N-terminal truncation (up to residue V161) with N-terminal fusion of an SV40 NLS, UL42 polypeptide, CTIP polypeptide, HMGN1 polypeptide; and C-terminal fusion of HCV core peptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, and A1282G.

[0947] SEQ ID NO: 1248 (encoded by SEQ ID NO: 11248) sets forth the amino acid sequence of a R2-TG RTE having an N-terminal truncation (up to residue V161) with N-terminal fusion of a bipartite SV40 NLS, canine hepatitis virus core peptide, UL42 polypeptide, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, and A1282G.

[0948] SEQ ID NO:1246 (encoded by SEQ ID NO: 11246) sets forth the amino acid sequence of a R2-TG RTE having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, UL42 polypeptide, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, K1157A, I1219N, and A1282G.

[0949] FIGS. 16A and 16B are bar graphs showing results of integration assays using engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOs: 636, 639, 648-650, 782, 1243, 1246, 1248, 1255, 1256, 1259, 1261, 1266, 1267, 1837.

[0950] The dataset included the following controls: Lipofectamine-only and GFP reporters SEQ ID NO:175 and SEQ ID NO:581 as negative controls; and WT R2-TG control (SEQ ID NO: 1 + SEQ ID NO: 175) as positive control. Error bars indicate standard error of two replicates.

[0951] Most of engineered R2-TG drivers exhibited improved activity above that of wild-type R2-TG including drivers having an amino acid sequence forth in any one of SEQ ID NO: 636, SEQ ID NO: 639, SEQ ID NO: 648, SEQ ID NO: 649, SEQ ID NO: 650, SEQ ID NO: 1243, SEQ ID NO: 782, SEQ ID NO: 1267, SEQ ID NO: 1266, SEQ ID NO: 1261, SEQ ID NO: 1259, SEQ ID NO: 1256, SEQ ID NO: 1255, SEQ ID NO: 1248, SEQ ID NO: 1246. In particular, SEQ ID NO: 1266 (encoded by SEQ ID NO: 11266) which included the EIF4A polypeptide that has RNA helicase activity (SEQ ID NO: 585), as well as DNA repair recruitment peptide CTIP (SEQ ID NO: 61), chromatin binding domain HMGN1 (SEQ ID NO: 40), SV40 NLS, and further includes amino acid substitutions G257R, N363H, E399M, V539G, H423Q, A442P, D537Y, K547R, D555K, R669L, N889G, F973E, T1011A, Q1019R, D1133P, K1085R, R1177G, Y1242K, and A1282G exhibited the highest activity.

[0952] Relative to the engineered R2-TG driver set forth in SEQ ID NO: 636, the driver having an amino acid sequence set forth in SEQ ID NO: 648 included the G658R amino acid substitution, and SEQ ID NO: 649 (encoded by SEQ ID NO: 10649) included the G642R amino acid substitution, each of which improved activity by ~20%.

[0953] Relative to the engineered R2-TG driver set forth in SEQ ID NO: 782, the driver having an amino acid sequence set forth in SEQ ID NO: 1248 includes an addition of a bipartite SV40 NLS, followed by a canine hepatitis virus core peptide (SEQ ID NO: 789), together which improved activity approximately 2-fold.

[0954] These results indicate the R2 activity-enhancing effect of these amino acid substitutions and peptide additions.

Example 17

[0955] Driver constructs encoding engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOS: 618-636 described, hereinbelow were generated and tested with the GFP reporter construct having a nucleic acid sequence set forth in SEQ ID NO: 581.

[0956] SEQ ID NO: 618 (encoded by SEQ ID NO: 10618) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue E184) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide.

[0957] SEQ ID NO: 619 (encoded by SEQ ID NO: 10619) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue E184) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, E399M, D555K, D923K, T1011S, I1219N, and A1282G.

[0958] SEQ ID NO: 620 (encoded by SEQ ID NO: 10620) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, Sto7D polypeptide, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, and A1282G.

[0959] SEQ ID NO: 621 (encoded by SEQ ID NO: 10621) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, Sto7D polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, and A1282G.

[0960] SEQ ID NO: 622 (encoded by SEQ ID NO: 10622) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide, Sto7D polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, and A1282G.

[0961] SEQ ID NO: 623 (encoded by SEQ ID NO: 10623) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue E184) with N-terminal fusion of a SV40 NLS, Sto7D polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, K700D, C518Y, and T989L.

[0962] SEQ ID NO: 624 (encoded by SEQ ID NO: 10624) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, Sto7D polypeptide, CTIP polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, K700D, C518Y, and T989L.

[0963] SEQ ID NO: 625 (encoded by SEQ ID NO: 10625) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue E184) with N-terminal fusion of a SV40 NLS, Sto7D polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, K700D.

[0964] SEQ ID NO: 626 (encoded by SEQ ID NO: 10626) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, Sto7D polypeptide, CTIP polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, K700D, and C518Y.

[0965] SEQ ID NO: 627 (encoded by SEQ ID NO: 10627) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue E184) with N-terminal fusion of a SV40 NLS, Sto7D polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, K700D.

[0966] SEQ ID NO: 628 (encoded by SEQ ID NO: 10628) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, Sto7D polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K.

[0967] SEQ ID NO: 629 (encoded by SEQ ID NO: 10629) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, Sso7D polypeptide, CTIP polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, and A1282G. The DNA encoding the polypeptide RT domain is codon-optimized for high frequency codons.

[0968] SEQ ID NO: 630 (encoded by SEQ ID NO: 10630) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue E184) with N-terminal fusion of a SV40 NLS, HMGN1 polypeptide, Sto7D polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, and D923K.

[0969] SEQ ID NO: 631 (encoded by SEQ ID NO: 10631) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, Sto7D polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, and K700D.

[0970] SEQ ID NO: 632 (encoded by SEQ ID NO: 10632) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, Sto7D polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, and A1282G.

[0971] SEQ ID NO: 633 (encoded by SEQ ID NO: 10633) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue E184) with N-terminal fusion of a SV40 NLS, Sto7D polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, and A1282G.

[0972] SEQ ID NO: 634 (encoded by SEQ ID NO: 10634) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue E184) with N-terminal fusion of a SV40 NLS, HMGN1 polypeptide, Sto7D polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, and A1282G.

[0973] SEQ ID NO: 635 (encoded by SEQ ID NO: 10635) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, Sto7D polypeptide, CTIP polypeptide.

[0974] FIGS. 17A and 17B are bar graphs showing results of integration assays using engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOs: 618-636.

[0975] All of the engineered R2 drivers exhibited improved activity above that of WT R2-TG control (SEQ ID NO: 1+SEQ ID NO:175) set forth in any one of SEQ ID NO: 618, SEQ ID NO:619, SEQ ID NO:620, SEQ ID NO:621, SEQ ID NO:622, SEQ ID NO:623, SEQ ID NO:624, SEQ ID NO:625, SEQ ID NO:626, SEQ ID NO:627, SEQ ID NO:628, SEQ ID NO:629, SEQ ID NO:630, SEQ ID NO:631, SEQ ID NO:632, SEQ ID NO:633, SEQ ID NO:634, SEQ ID NO:635, or SEQ ID NO:636.

[0976] Six engineered R2-TG drivers exhibited improved activity over WT R2-TG (SEQ ID NO:1). SEQ ID NO:636 with CTIP and HMGN1 domains exhibited ~2x fold improvement activity over WT R2-TG and five others had improved activity over SEQ ID NO:636 with ~3.5x fold improvement activity over WT. SEQ ID NO:623 (encoded by SEQ ID NO: 10623) has a longer truncation than SEQ ID NO:636 and includes Sto7D N-terminal domain fusion and three additional amino acid substitutions: K700D, C518Y and T989L, SEQ ID NO:625 with HMGN1 and Sto7D in reverse order compared to SEQ ID NO:623, SEQ ID NO:623 with a K700D amino acid substitution compared to SEQ ID NO:636, SEQ ID NO:627, SEQ ID NO:627 is similar to SEQ ID NO:623, SEQ ID NO:623 with a K700D amino acid substitution compared to SEQ ID NO:636; SEQ ID NO:633 is similar to SEQ ID NO:623 and with the same

amino acid substitutions as in SEQ ID NO:636. SEQ ID NO:634 is similar to SEQ ID NO:633 with Sto7D replacing HMGN1.

Example 18: Amino acid substitutions in the R2-TG polypeptide contributes to improved activity

[0977] Driver constructs encoding engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOS: 637-650 described, hereinbelow were generated and tested with the GFP reporter construct having a nucleic acid sequence set forth in SEQ ID NO: 581.

[0978] SEQ ID NO:637 (encoded by SEQ ID NO: 10637) sets forth the amino acid sequence of an engineered R2-TG driver polypeptide with a H1240A amino acid substitution.

[0979] SEQ ID NO:638 (encoded by SEQ ID NO: 10638) sets forth the amino acid sequence of an engineered R2-TG driver polypeptide with a Y1131A amino acid substitution.

[0980] SEQ ID NO:639 (encoded by SEQ ID NO: 10639) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, and A1282G. In this sequence, residues 610-621 in the R2-TG protein sequence were replaced by TEIQTTIREYYK residues.

[0981] SEQ ID NO:640 (encoded by SEQ ID NO: 10640) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K and A877V.

[0982] SEQ ID NO:641 (encoded by SEQ ID NO: 10641) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K and A877I.

[0983] SEQ ID NO:642 (encoded by SEQ ID NO: 10642) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K and A877M.

[0984] SEQ ID NO:643 (encoded by SEQ ID NO: 10643) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K and F876Y.

[0985] SEQ ID NO:644 (encoded by SEQ ID NO: 10644) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K and A877H.

[0986] SEQ ID NO:645 (encoded by SEQ ID NO: 10645) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K and Q955R.

[0987] SEQ ID NO:646 (encoded by SEQ ID NO: 10646) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K and L952I. The nucleic acid sequence encoding the polypeptide is set forth in SEQ ID NO: 10646.

[0988] SEQ ID NO:647 (encoded by SEQ ID NO: 10647) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K and E947S.

[0989] SEQ ID NO:648 (encoded by SEQ ID NO: 10648) is described hereinabove.

[0990] SEQ ID NO:649 (encoded by SEQ ID NO: 10649) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K and Q642R.

[0991] SEQ ID NO:650 (encoded by SEQ ID NO: 10650) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal

fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K and F973R.

[0992] FIGS. 18A and 18B are bar graphs showing results of integration assays using engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOs: 636-650.

[0993] Most of the engineered R2-TG drivers exhibited improved activity above that of wild-type R2-TG (SEQ ID NO:1): SEQ ID NO:639, SEQ ID NO:643, SEQ ID NO:645, SEQ ID NO:646, SEQ ID NO:647, SEQ ID NO:648, SEQ ID NO:649, SEQ ID NO:650, SEQ ID NO:636.

[0994] SEQ ID NO:636 exhibited a ~2.25x fold improvement activity over WT and six others had activity higher than SEQ ID NO: 636 with ~3x fold improvement activity over WT (SEQ ID NO:1). Without wishing to be bound to theory, SEQ ID NO:645 with positively a charged Q955R amino acid substitution, and SEQ ID NO:646 with the L952I amino acid substitution may have stabilized the modified driver. The effect of the other amino acid substitutions may be related to stabilization, reduced steric hindrance and or increased affinity to RNA. For example, SEQ ID NO:647 with a E947S amino acid substitution, SEQ ID NO:648 with the G658R amino acid substitution, SEQ ID NO:649 with positively charged Q642R amino acid substitution, and SEQ ID NO:650 with positively charged F973E amino acid substitution.

Example 19

[0995] Driver constructs encoding engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOS: 651-668 were generated and tested with the GFP reporter construct having a nucleic acid sequence set forth in SEQ ID NO: 581.

[0996] SEQ ID NO:651 (encoded by SEQ ID NO: 10651) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and D227W.

[0997] SEQ ID NO:652 (encoded by SEQ ID NO: 10652) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal

fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and C344V.

[0998] SEQ ID NO:653 (encoded by SEQ ID NO: 10653) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and R669V.

[0999] SEQ ID NO:654 (encoded by SEQ ID NO: 10654) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and T989V.

[1000] SEQ ID NO:655 (encoded by SEQ ID NO: 10655) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and G689E.

[1001] SEQ ID NO:656 (encoded by SEQ ID NO: 10656) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and T679K.

[1002] SEQ ID NO: 657 (encoded by SEQ ID NO: 10657) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and F956V.

[1003] SEQ ID NO: 658 (encoded by SEQ ID NO: 10658) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and M686S.

[1004] SEQ ID NO:659 (encoded by SEQ ID NO: 10659) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal

fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and W1126P.

[1005] SEQ ID NO:660 (encoded by SEQ ID NO: 10660) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and C404S.

[1006] SEQ ID NO:661 (encoded by SEQ ID NO: 10661) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and I49C amino acid substitution in the CTIP polypeptide.

[1007] SEQ ID NO:662 (encoded by SEQ ID NO: 10662) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and I2A amino acid substitution in the CTIP polypeptide.

[1008] SEQ ID NO:663 (encoded by SEQ ID NO: 10663) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K and D1133N.

[1009] SEQ ID NO:664 (encoded by SEQ ID NO: 10664) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K and N941K.

[1010] SEQ ID NO:665 (encoded by SEQ ID NO: 10665) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K and A729K.

[1011] SEQ ID NO:666 (encoded by SEQ ID NO: 10666) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K and M874L.

[1012] SEQ ID NO:667 (encoded by SEQ ID NO: 10667) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K, A649S and M850L.

[1013] SEQ ID NO:668 (encoded by SEQ ID NO: 10668) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, EIF4a polypeptide, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions G257R, P280K, N363H, E399M, V539G, H423Q, A442P, D537Y, K547R, D555K, R669L, F973E, T1011A, Q1019R, D1133P, K1085R, R1177G, Y1242K, A1282G.

[1014] FIGS. 19A and 19B are bar graphs showing results of integration assays using engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOs: 651-668.

[1015] All of the engineered R2-TG driver polypeptides exhibited improved activity above that of WT R2-TG control. SEQ ID NO: 636 exhibited ~3x fold improvement activity over WT R2-TG. SEQ ID NO:654 with a T989V stabilizing amino acid substitution and SEQ ID NO:668 (with several amino acid substitutions had more than 6x improvement over WT R2-TG driver and GFP reporter control (SEQ ID NO:1 + SEQ ID NO:175).

Example 20: R2 RTE comprising a D923K amino acid substitution show improved activity.

[1016] Driver constructs encoding engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOS: 669- 686 were generated and tested with the GFP reporter construct having a nucleic acid sequence set forth in SEQ ID NO: 581.

[1017] SEQ ID NO:669 (encoded by SEQ ID NO: 10669) sets forth the amino acid sequence of a R2-TG with a TEV 5' UTR in place of the native R2-TG 5' RTE-UTR, N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1

polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, and A1282G.

[1018] SEQ ID NO:670 (encoded by SEQ ID NO: 10670) sets forth the amino acid sequence of a R2-TG with N-terminal truncation (up to residue E218).

[1019] SEQ ID NO:671 (encoded by SEQ ID NO: 10671) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue E218) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, E399M, D555K, T1011S, I1219N, A1282G and D923K.

[1020] SEQ ID NO:672 (encoded by SEQ ID NO: 10672) sets forth the amino acid sequence of a R2-TG with R1337A amino acid substitution. The nucleic acid sequence encoding the polypeptide is set forth in SEQ ID NO: 10672.

[1021] SEQ ID NO:673 (encoded by SEQ ID NO: 10673) sets forth the amino acid sequence of a R2-TG with H1240A and R1337A amino acid substitutions.

[1022] SEQ ID NO:674 (encoded by SEQ ID NO: 10674) sets forth the amino acid sequence of a R2-TG with H1240A and Y1311A amino acid substitutions. The nucleic acid sequence encoding the polypeptide is set forth in SEQ ID NO: 10674.

[1023] SEQ ID NO:675 (encoded by SEQ ID NO: 10675) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue E184) with N-terminal fusion of a SV40 NLS, UL42 polypeptide, CTIP polypeptide, HMGN1 polypeptide; C-terminal fusion of CHV core peptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and D923K.

[1024] SEQ ID NO:676 (encoded by SEQ ID NO: 10676) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G. In this sequence, residues 613-621 in the R2-TG protein sequence were replaced by QTTIREYYK residues.

[1025] SEQ ID NO:677 (encoded by SEQ ID NO: 10677) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal

fusion of a MYC NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G.

[1026] SEQ ID NO:678 (encoded by SEQ ID NO: 10678) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, TUDOR polypeptide, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G.

[1027] SEQ ID NO:679 (encoded by SEQ ID NO: 10679) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, hnRNPA1a peptide, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K.

[1028] SEQ ID NO:680 (encoded by SEQ ID NO: 10680) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, hnRNPA1b peptide, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K.

[1029] SEQ ID NO:681 (encoded by SEQ ID NO: 10681) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, DDX4a peptide, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K.

[1030] SEQ ID NO:682 (encoded by SEQ ID NO: 10682) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CWC27b peptide, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K.

[1031] SEQ ID NO: 683 (encoded by SEQ ID NO: 10683) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, SPAG7a peptide, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K.

[1032] SEQ ID NO:684 (encoded by SEQ ID NO: 10684) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal

fusion of a SV40 NLS, RS4Y2a peptide, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K.

[1033] SEQ ID NO:685 (encoded by SEQ ID NO: 10685) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, H2A1Ha peptide, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K.

[1034] SEQ ID NO:686 (encoded by SEQ ID NO: 10686) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, RAMACa peptide, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K.

[1035] SEQ ID NO: 636 (encoded by SEQ ID NO: 10636) is described herein, supra.

[1036] FIGS. 20A and 20B are bar graphs showing results of integration assays using engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOS: 669-686.

[1037] Driver constructs encoding engineered R2-TG drivers having amino acid sequences set forth in any one of SEQ ID NO: 669, 670, 671, 673, 675, 676, 677, 678, 679, 680, 681, 682, 683, 684, 685, 686 exhibited improved activity above that of wild-type R2-TG. SEQ ID NO:636 (encoded by SEQ ID NO: 10636) with CTIP and HMGN1 domains but without a D923K amino acid substitution had ~4x fold improvement activity over WT control (SEQ ID NO:1+SEQ ID NO:175). SEQ ID NO: 678, 679, 680, 681, 682 and 683 exhibited improved activity over the SEQ ID NO:636 system control. SEQ ID NO: 678 includes an additional RNA helicase Tudor domain. Each of SEQ ID NOS: 682, 683, 684, 685 and 686 have the D923K amino acid substitution and an additional heterologous fusion of a disordered region of condensation proteins.

Example 21

[1038] Driver constructs encoding engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOS: 687-703 were generated and tested with the GFP reporter construct having a nucleic acid sequence set forth in SEQ ID NO: 581.

[1039] SEQ ID NO:687 (encoded by SEQ ID NO: 10687) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, the first 184 WT R2-TG peptide, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K.

[1040] SEQ ID NO:688 (encoded by SEQ ID NO: 10688) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, the first 218 WT R2-TG peptide, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K.

[1041] SEQ ID NO:689 (encoded by SEQ ID NO: 10689) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, UBQ2a, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K.

[1042] SEQ ID NO:690 (encoded by SEQ ID NO: 10690) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, UBQ2b, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K.

[1043] SEQ ID NO:691 (encoded by SEQ ID NO: 10691) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, RBM14b, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K.

[1044] SEQ ID NO:692 (encoded by SEQ ID NO: 10692) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, RBM14d, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K.

[1045] SEQ ID NO:693 (encoded by SEQ ID NO: 10693) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal

fusion of a SV40 NLS, TAF15c, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K.

[1046] SEQ ID NO:694 (encoded by SEQ ID NO: 10694) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, TAF15d, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K.

[1047] SEQ ID NO:695 (encoded by SEQ ID NO: 10695) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, TDP43a, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K.

[1048] SEQ ID NO:696 (encoded by SEQ ID NO: 10696) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, TDP43b, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K.

[1049] SEQ ID NO:697 (encoded by SEQ ID NO: 10697) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, FUSb, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K.

[1050] SEQ ID NO:698 (encoded by SEQ ID NO: 10698) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CWC27a, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K.

[1051] SEQ ID NO:699 (encoded by SEQ ID NO: 10699) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, SPAG7b, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K.

[1052] SEQ ID NO:700 (encoded by SEQ ID NO: 10700) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal

fusion of a SV40 NLS, RBMY1Da, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K.

[1053] SEQ ID NO:701 (encoded by SEQ ID NO: 10701) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, RBMY1Db, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K.

[1054] SEQ ID NO:702 (encoded by SEQ ID NO: 10702) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, KHDC4b, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K.

[1055] SEQ ID NO:703 (encoded by SEQ ID NO: 10703) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, KHDC4c, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K.

[1056] FIGS. 21A and 21B are bar graphs showing results of integration assays using the R2-TG drivers having amino acid sequences set forth in SEQ ID NOs: 687-703.

[1057] All the engineered R2-TG drivers exhibited improved activity above that of wild-type R2-TG (SEQ ID NO:1). Seven modifications had improved activity over SEQ ID NO:636 (encoded by SEQ ID NO: 10636), which was used as another positive control: SEQ ID NO:689 SEQ ID NO:696 SEQ ID NO:697 SEQ ID NO:699 SEQ ID NO:701 SEQ ID NO:702 SEQ ID NO:703 The engineered R2-TG drivers with improved activity have the D923K substitution within the SEQ ID NO:636 (encoded by SEQ ID NO: 10636) and also additional peptides of disordered regions of proteins that tend to condensate.

Example 22: Presence of condensation protein fusions improve R2 RTE activity

[1058] Driver constructs encoding engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOS: 705-721 and 1687 were generated and tested with the GFP reporter construct having a nucleic acid sequence set forth in SEQ ID NO: 581.

[1059] SEQ ID NO:705 (encoded by SEQ ID NO: 10705) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, RP9a, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K.

[1060] SEQ ID NO:706 (encoded by SEQ ID NO: 10706) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, RP9b, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K.

[1061] SEQ ID NO:707 (encoded by SEQ ID NO: 10707) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, PHP14a, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K.

[1062] SEQ ID NO:708 (encoded by SEQ ID NO: 10708) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CWC25a, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K.

[1063] SEQ ID NO:709 (encoded by SEQ ID NO: 10709) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, RAD51AP1a, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K.

[1064] SEQ ID NO:710 (encoded by SEQ ID NO: 10710) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, SPA24a, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K.

[1065] SEQ ID NO:711 (encoded by SEQ ID NO: 10711) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, H1Ta, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K.

[1066] SEQ ID NO:712 (encoded by SEQ ID NO: 10712) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, AIMP1a, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K.

[1067] SEQ ID NO:713 (encoded by SEQ ID NO: 10713) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, MRPL1a, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K.

[1068] SEQ ID NO:714 (encoded by SEQ ID NO: 10714) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, the first 161 WT R2-TG peptide, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K.

[1069] SEQ ID NO:715 (encoded by SEQ ID NO: 10715) is described hereinabove.

[1070] SEQ ID NO:716 (encoded by SEQ ID NO: 10716) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a bipSV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G.

[1071] SEQ ID NO:717 (encoded by SEQ ID NO: 10717) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a NOA36 NoLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G.

[1072] SEQ ID NO:718 (encoded by SEQ ID NO: 10718) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40r6 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G.

[1073] SEQ ID NO:719 (encoded by SEQ ID NO: 10719) sets forth the amino acid sequence of a R2-TG polypeptide that includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N. The DNA encoding the polypeptide is codon optimized for high frequency codons.

[1074] SEQ ID NO:720 (encoded by SEQ ID NO: 10720) sets forth the amino acid sequence of a R2-TG polypeptide that includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G. The DNA encoding the polypeptide is codon optimized for high frequency codons.

[1075] SEQ ID NO:721 (encoded by SEQ ID NO: 10721) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G. The DNA encoding the R2-TG polypeptide is codon optimized for high frequency codons.

[1076] SEQ ID NO:1687 (encoded by SEQ ID NO: 11687) sets forth the amino acid sequence of an R2-TG polypeptide. R2-TG polypeptide includes amino acid substitutions P280K, D555K, T1011S, I1219N. The DNA encoding the polypeptide is codon optimized for high frequency codons.

[1077] FIGS. 22A and 22B are bar graphs showing results of integration assays using engineered R2-TG drivers having an amino acid sequence set forth in SEQ ID NOs: 705-721 and a GFP reporter having nucleic acid sequence set forth in SEQ ID NO:581.

[1078] All the engineered R2-TG drivers exhibited improved activity above that of wild-type R2-TG. Several drivers exhibited improved activity over SEQ ID NO:636 (encoded by SEQ ID NO: 10636) including SEQ ID NO:708, SEQ ID NO:712, SEQ ID NO:713, SEQ ID NO:714, SEQ ID NO:715, SEQ ID NO:716, SEQ ID NO:717, SEQ ID NO:718, SEQ ID NO:719, SEQ ID NO:720, SEQ ID NO:721, SEQ ID NO:1687.

[1079] Drivers with improved activity have the D923K amino acid substitution in addition to the modifications present in SEQ ID NO:636 (encoded by SEQ ID NO: 10636) and further include peptide fusion of disordered regions of proteins that tend to condensate. The most active drivers are SEQ ID NO:708, SEQ ID NO:712, SEQ ID NO:713, SEQ ID NO:714. These drivers exhibited a 1.3-2.5 fold improved activity over engineered R2-TG and modified reporter (SEQ

ID NO:636 +SEQ ID NO:581). Drivers with the D923K amino acid substitution and modified NLS designs yielded 1.8-2x fold improvement over engineered R2-TG and modified reporter (SEQ ID NO:636 +SEQ ID NO:581 and include SEQ ID NO:716, SEQ ID NO:717 and SEQ ID NO:718. Codon optimized designs improved activity 2-3 fold over engineered R2-TG and modified reporter (SEQ ID NO:636 +SEQ ID NO:581) including SEQ ID NO:1687, SEQ ID NO:719, SEQ ID NO:720 and SEQ ID NO:721.

Example 23

[1080] Driver constructs encoding engineered R2-TG drivers having amino acid sequences set forth in an SEQ ID NOS: 722-739 were generated to test the effect of N-terminal truncated R2-TG together with a variety of amino acid substitutions in the truncated R2-TG and tested with the GFP reporter construct having a nucleic acid sequence set forth in SEQ ID NO: 581.

[1081] SEQ ID NO:722 (encoded by SEQ ID NO: 10722) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K and G647T.

[1082] SEQ ID NO:723 (encoded by SEQ ID NO: 10723) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K and D990V.

[1083] SEQ ID NO:724 (encoded by SEQ ID NO: 10724) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K and D990R.

[1084] SEQ ID NO:725 (encoded by SEQ ID NO: 10725) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K and V727L.

[1085] SEQ ID NO:726 (encoded by SEQ ID NO: 10726) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K, Y571H, S645K, M850L.

[1086] SEQ ID NO:727 (encoded by SEQ ID NO: 10727) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K, Y571H, S645K, M841L, M850L.

[1087] SEQ ID NO:728 (encoded by SEQ ID NO: 10728) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, E399M, D555K, T1011S, I1219N, A1282G, D923K, Y571H, S645K, A649S, M850L.

[1088] SEQ ID NO:729 (encoded by SEQ ID NO: 10729) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K and D227F.

[1089] SEQ ID NO:730 (encoded by SEQ ID NO: 10730) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K and V600T.

[1090] SEQ ID NO:731 (encoded by SEQ ID NO: 10731) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K and M449E.

[1091] SEQ ID NO:732 (encoded by SEQ ID NO: 10732) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal

fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K and Y478L.

[1092] SEQ ID NO:733 (encoded by SEQ ID NO: 10733) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K and M671S.

[1093] SEQ ID NO:734 (encoded by SEQ ID NO: 10734) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K and R669F.

[1094] SEQ ID NO:735 (encoded by SEQ ID NO: 10735) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K and M420T.

[1095] SEQ ID NO:736 (encoded by SEQ ID NO: 10736) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K and T1367R.

[1096] SEQ ID NO:737 (encoded by SEQ ID NO: 10737) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and S17D amino acid substitution in the CTIP polypeptide.

[1097] SEQ ID NO:738 (encoded by SEQ ID NO: 10738) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and D18C amino acid substitution in the CTIP polypeptide.

[1098] SEQ ID NO:739 (encoded by SEQ ID NO: 10739) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and Q46T amino acid substitution in the CTIP polypeptide.

[1099] FIGS. 23A and 23B are bar graphs showing results of integration assays using engineered R2-TG drivers having amino acid sequences in SEQ ID NOs: 722-739. The R2-TG drivers were used with the GFP reporter construct SEQ ID NO:581.

[1100] All the engineered R2-TG drivers exhibited improved activity above that of wild-type R2-TG control (SEQ ID NO:1, SEQ ID NO:125). All samples except SEQ ID NO:723 increased activity over SEQ ID NO:636 by ~2-3-fold. The following six engineered R2-TG drivers exhibited about 3-fold improved activity: SEQ ID NO:722, SEQ ID NO:730, SEQ ID NO:736, SEQ ID NO:737, SEQ ID NO:738 and SEQ ID NO:739.

Example 24

[1101] The following engineered R2-TG drivers having an amino acid sequence set forth in an SEQ ID NOS: 740-757 were generated to test the effect of N-terminal truncated R2-TG together with a variety of amino acid substitutions in the truncated R2-TG and were tested with the GFP reporter construct having a nucleic acid sequence set forth in SEQ ID NO: 581

[1102] SEQ ID NO:740 (encoded by SEQ ID NO: 10740) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and C198S amino acid substitution in the CTIP polypeptide.

[1103] SEQ ID NO:741 (encoded by SEQ ID NO: 10741) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and E155S amino acid substitution in the CTIP polypeptide.

[1104] SEQ ID NO:742 (encoded by SEQ ID NO: 10742) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and L281G amino acid substitution in the CTIP polypeptide.

[1105] SEQ ID NO:743 (encoded by SEQ ID NO: 10743) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and I186S amino acid substitution in the CTIP polypeptide.

[1106] SEQ ID NO:744 (encoded by SEQ ID NO: 10744) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and Q46K amino acid substitution in the CTIP polypeptide.

[1107] SEQ ID NO:745 (encoded by SEQ ID NO: 10745) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and D18F amino acid substitution in the CTIP polypeptide.

[1108] SEQ ID NO:746 (encoded by SEQ ID NO: 10746) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K and E1363K.

[1109] SEQ ID NO:747 (encoded by SEQ ID NO: 10747) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K and Q642R.

[1110] SEQ ID NO:748 (encoded by SEQ ID NO: 10748) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K and V1368R.

[1111] SEQ ID NO:749 (encoded by SEQ ID NO: 10749) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K and R1157M.

[1112] SEQ ID NO:750 (encoded by SEQ ID NO: 10750) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K and N1123K.

[1113] SEQ ID NO:751 (encoded by SEQ ID NO: 10751) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K and P1116D.

[1114] SEQ ID NO:752 (encoded by SEQ ID NO: 10752) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K and D879A.

[1115] SEQ ID NO:753 (encoded by SEQ ID NO: 10753) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K and P998H.

[1116] SEQ ID NO:754 (encoded by SEQ ID NO: 10754) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K and D879M.

[1117] SEQ ID NO:755 (encoded by SEQ ID NO: 10755) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K and S645Q.

[1118] SEQ ID NO:756 (encoded by SEQ ID NO: 10756) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K and P998Q.

[1119] SEQ ID NO:757 (encoded by SEQ ID NO: 10757) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K and F846Y.

[1120] FIGS. 24A and 24B are bar graphs showing results of integration assays using engineered R2-TG drivers having an amino acid sequence set forth in SEQ ID NOs: 740-757. Most of the engineered R2-TG drivers exhibited improved activity above that of wild-type R2-TG (SEQ ID NO:1 +SEQ ID NO:175). Several engineered drivers with modifications described herein exhibited improved activity over SEQ ID NO:636. Six engineered drivers with amino acid substitutions in the CTIP polypeptide improve by 1.25-1.4 over the control engineered R2-TG driver (SEQ ID NO:636 +SEQ ID NO:581) and include SEQ ID NO:740, SEQ ID NO:741, SEQ ID NO:742, SEQ ID NO:743, SEQ ID NO:744, and SEQ ID NO:745.

[1121] The following nine drivers exhibited improved activity over SEQ ID NO:636 with up to 1.75 fold improvement: SEQ ID NO:746, SEQ ID NO:747, SEQ ID NO:748, SEQ ID NO:749, SEQ ID NO:750, SEQ ID NO:751, SEQ ID NO:755, SEQ ID NO:756 and SEQ ID NO:757.

[1122] Two drivers exhibited reduced activity to below the WT levels: SEQ ID NO: 752 and SEQ ID NO:754.

Example 25

[1123] The engineered R2-TG driver polypeptides having an amino acid sequence set forth in SEQ ID NOS: 760--774, described hereinbelow were generated and tested. In this experiment,

different retrotransposable mutants and reconstructed elements driver constructs were used with the GFP reporter construct (SEQ ID NO: 581)

[1124] SEQ ID NO: 760 (encoded by SEQ ID NO: 10760) sets forth the amino acid sequence of a R2-TG N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide. R2-TG polypeptide includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and L848I as well as the HAG 5'UTR, HAG 3'UTR and 3 stop codons.

[1125] SEQ ID NO: 761 (encoded by SEQ ID NO: 10761) sets forth the amino acid sequence of a R2-TG N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes the amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and L854I as well as the HAG 5'UTR, HAG 3'UTR and three stop codons.

[1126] SEQ ID NO: 762 (encoded by SEQ ID NO: 10762) sets forth the amino acid sequence of a R2-TG N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes the amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and F1162Y as well as the HAG 5'UTR, HAG 3'UTR and 3 stop codons.

[1127] SEQ ID NO: 763 (encoded by SEQ ID NO: 10763) sets forth the amino acid sequence of a R2-TG N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes the amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and Q1173R as well as the HAG 5'UTR, HAG 3'UTR and 3 stop codons.

[1128] SEQ ID NO: 764 (encoded by SEQ ID NO: 10764) sets forth the amino acid sequence of a R2 created by ancestral reconstruction process on using R2 elements. The R2 element was reconstructed from monophyletic group 1 node 7 whose leaves are R2-MC (SEQ ID NO:411), R2-MC (SEQ ID NO:458), R2-GA (SEQ ID NO: 408), R2-GA (SEQ ID NO: 455), R2-HN (SEQ ID NO:409), R2-HN (SEQ ID NO:456) as well as the HAG 5'UTR, HAG 3'UTR and 3 stop codons. The DNA encoding the polypeptide is codon-optimized using GenSmart™ from GenScript.

[1129] SEQ ID NO: 770 (encoded by SEQ ID NO: 10770) sets forth the amino acid sequence of a R2 created by ancestral reconstruction process on using R2 elements. The R2 element was reconstructed from monophyletic group 1 node 8 whose leaves are R2-GA (SEQ ID NO:408), R2-GA (SEQ ID NO:455), R2-HN (SEQ ID NO:409), R2-HN (SEQ ID NO:456) as well as the HAG 5'UTR, HAG 3'UTR and 3 stop codons. The DNA encoding the polypeptide is codon-optimized using GenSmart™ from GenScript.

[1130] SEQ ID NO: 771 (encoded by SEQ ID NO: 10771) sets forth the amino acid sequence of a R2 created by ancestral reconstruction process on using R2 elements. The R2 element was reconstructed from monophyletic group 2 node 9 whose leaves are R2-XC (SEQ ID NO:474), R2-LS (SEQ ID NO: 457), R2-LS (SEQ ID NO:410), R2-GP (SEQ ID NO: 453), R2-GP (SEQ ID NO:407), R2-SM (SEQ ID NO:480), and R2-SM (SEQ ID NOS:462) as well as the HAG 5'UTR, HAG 3'UTR and 3 stop codons. The DNA encoding the polypeptide is codon-optimized using GenSmart™ from GenScript.

[1131] SEQ ID NO: 774 (encoded by SEQ ID NO: 10774) sets forth the amino acid sequence of a R2 created by ancestral reconstruction process on using R2 elements. The R2 element was reconstructed from monophyletic group 2 node 10 whose leaves are R2-LS (SEQ ID NO: 457) and R2-LS (SEQ ID NOS: 410) as well as the HAG 5'UTR, HAG 3'UTR and 3 stop codons. The DNA encoding the polypeptide is codon-optimized using GenSmart™ from GenScript.

[1132] SEQ ID NO: 769 (encoded by SEQ ID NO: 10769) sets forth the amino acid sequence of a R2 created by ancestral reconstruction process on using R2 elements. The R2 element was reconstructed from monophyletic group 2 node 11 whose leaves are R2-GP (SEQ ID NO: 453), R2-GP (SEQ ID NO:407), R2-SM (SEQ ID NO:480), and R2-SM (SEQ ID NO:462) as well as the HAG 5'UTR, HAG 3'UTR and 3 stop codons. The DNA encoding the polypeptide is codon-optimized using GenSmart™ from GenScript.

[1133] SEQ ID NO: 772 (encoded by SEQ ID NO: 10772) sets forth the amino acid sequence of a R2 created by ancestral reconstruction process on using R2 elements. The R2 element was reconstructed from monophyletic group 6 node 10 whose leaves are R2-HN SEQ ID NO: 409), R2-HN (SEQ ID NO: 456), R2-MC (SEQ ID NO: 458), R2-MC (SEQ ID NO:411), R2-OT (SEQ ID NO: 460), R2-OT (SEQ ID NO:1336), R2-OT (SEQ ID NO:413), R2-GA(SEQ ID NO: 455) and R2-GA (SEQ ID NO: 408) as well as the HAG 5'UTR, HAG 3'UTR and 3 stop

codons. The DNA encoding the polypeptide is codon-optimized using GenSmart™ from GenScript.

[1134] SEQ ID NO: 773 (encoded by SEQ ID NO: 10773) sets forth the amino acid sequence of a R2 created by ancestral reconstruction process on using R2 elements. The R2 element was reconstructed from monophyletic group 6 node 13 whose leaves are R2-OT (SEQ ID NO: 460), R2-OT (SEQ ID NO:1336), R2-OT (SEQ ID NO:413), R2-GA (SEQ ID NO:455) and R2-GA (SEQ ID NO:408) as well as the HAG 5'UTR, HAG 3'UTR and 3 stop codons. The DNA encoding the polypeptide is codon-optimized using GenSmart™ from GenScript.

[1135] SEQ ID NO: 765 (encoded by SEQ ID NO: 10765) sets forth the amino acid sequence of a R2 created by ancestral reconstruction process on using R2 elements. The R2 element was reconstructed from monophyletic group 6 node 14 whose leaves are R2-OT (SEQ ID NO: 460), R2-OT (SEQ ID NO:1336) and R2-OT (SEQ ID NO:413) as well as the HAG 5'UTR, HAG 3'UTR and 3 stop codons. The DNA encoding the polypeptide is codon-optimized using GenSmart™ from GenScript.

[1136] SEQ ID NO: 766 (encoded by SEQ ID NO: 10766) sets forth the amino acid sequence of a R2-TG variant #2 found in R2-TG genome as well as the HAG 5'UTR, HAG 3'UTR and 3 stop codons. The DNA encoding the polypeptide is codon-optimized using fastcodon.

[1137] SEQ ID NO: 767 (encoded by SEQ ID NO: 10767) sets forth the amino acid sequence of a R2-TG variant #4 found in R2-TG genome as well as the HAG 5'UTR, HAG 3'UTR and 3 stop codons. The DNA encoding the polypeptide is codon-optimized using fastcodon.

[1138] SEQ ID NO: 768 (encoded by SEQ ID NO: 10768) sets forth the amino acid sequence of a R2-chimera with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide generated by R2-LS (SEQ ID NO:410) (residues 73-207)+ R2-TG (SEQ ID NO:636) (residues 311-433) + R2-GP (SEQ ID NO: 453) (residues 290-1139) + R2-TG (SEQ ID NO:636) (residues 1291-1390) as well as the HAG 5'UTR, HAG 3'UTR and 3 stop codons.

[1139] FIGS. 25A and 25B is a bar graph showing results of integration assays using drivers comprising R2-TG single mutants and R2 retrotransposable elements constructed by ancestral reconstruction process of active R2 elements (SEQ ID NOs: 760-767, 769-774) tested with engineered R2 reporter (SEQ ID NO: 581), a retrotransposable chimera element (SEQ ID NO:

768) composed from R2-TG, R2-LS and R2-GP (SEQ ID NOS: 636, 410, and 453). The R2-chimera was tested with either the engineered GFP reporter construct (SEQ ID NO: 581), R2-LS (SEQ ID NO: 758) or R2-GP (SEQ ID NO: 759). Transfection was carried out with driver and reporter mRNA delivered with Lipofectamine™ to HUH7 cells and full-length integration assessed by FACS at 24h.

[1140] The drivers were R2-TG amino acid substitutions, a R2-chimera protein constructs built from R2-TG, R2-LS and R2-GP and R2 constructs created by the ancestral reconstruction process using active R2 elements.

[1141] Most of the engineered drivers exhibited improved activity above that of WT control (SEQ ID NO: 1 + SEQ ID NO: 175): SEQ ID NO:760, SEQ ID NO:761, SEQ ID NO:762, SEQ ID NO:763, SEQ ID NO:764, SEQ ID NO: 770, SEQ ID NO: 771, SEQ ID NO: 774, SEQ ID NO: 769, SEQ ID NO: 772, SEQ ID NO: 773, SEQ ID NO: 765 and SEQ ID NO: 768.

Engineered R2-TG with CTIP and HMGN1 domains had ~5x fold improvement activity over the WT control. Two single mutant constructs had improved activity over SEQ ID NO:636 with 6-7x fold improvement activity over WT control. SEQ ID NO: 770 with an N-terminal truncation (up to residue V161), N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and L848I and SEQ ID NO: 763 with an N-terminal truncation (up to residue V161), N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and Q1173R.

Example 26

[1142] The following engineered R2-TG drivers having an amino acid sequence set forth in SEQ ID NOS: 775-784 were generated and tested with the GFP reporter construct having a nucleic acid sequence set forth in SEQ ID NO: 581 to test the effect of N-terminal truncated R2-TG with condensation peptide fusions and together with a variety of amino acid substitutions in the truncated R2-TG.

[1143] SEQ ID NO:775 (encoded by SEQ ID NO: 10775) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal

fusion of a SV40 NLS, UBQ2c peptide, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K.

[1144] SEQ ID NO:776 (encoded by SEQ ID NO: 10776) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, FUSc peptide, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K.

[1145] SEQ ID NO:777 (encoded by SEQ ID NO: 10777) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, DYRK3a peptide, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K.

[1146] SEQ ID NO:778 (encoded by SEQ ID NO: 10778) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, RS10La peptide, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K.

[1147] SEQ ID NO:779 (encoded by SEQ ID NO: 10779) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, LMOD1a peptide, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K.

[1148] SEQ ID NO:780 (encoded by SEQ ID NO: 10780) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, LMOD1b peptide, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K.

[1149] SEQ ID NO:781 (encoded by SEQ ID NO: 10781) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, RAD51AP1b peptide, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K.

[1150] SEQ ID NO:782 (encoded by SEQ ID NO: 10782) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal

fusion of a SV40 NLS, UL42 polypeptide, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G.

[1151] SEQ ID NO:783 (encoded by SEQ ID NO: 10783) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, EIF4a polypeptide, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G.

[1152] SEQ ID NO:784 (encoded by SEQ ID NO: 10784) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and Y571H.

[1153] SEQ ID NO:715 (encoded by SEQ ID NO: 10715) is described hereinabove.

[1154] FIGS. 26A and 26B are bar graphs showing results of integration assays using engineered R2-TG drivers having an amino acid sequence set forth in SEQ ID NOs: 775-784. The engineered R2-TG drivers were used with the GFP reporter construct SEQ ID NO:581.

[1155] All the engineered R2-TG drivers described in this example exhibited improved activity above that of WT R2-TG control (SEQ ID NO:1 +SEQ ID NO:175). Several engineered drivers with modifications described herein exhibited improved activity over the engineered R2-TG driver and reporter (SEQ ID NO:636 +SEQ ID NO:581): SEQ ID NO:776 (, SEQ ID NO:778, SEQ ID NO:779, SEQ ID NO:780 SEQ ID NO:781 . Three engineered drivers with peptide condensation peptides fusions also had higher activity than a similar construct with D923K amino acid substitution but without the addition of the condensation peptide (SEQ ID NO:715): SEQ ID NO:778, SEQ ID NO:779 and SEQ ID NO:780. The additional improved constructs that are comparable to SEQ ID NO:636 are SEQ ID NO:782 with a UL42 domain (SEQ ID NO: 587) addition over SEQ ID NO:636, SEQ ID NO:783 with an EIF4a domain addition over SEQ ID NO:636, SEQ ID NO:784 with a Y571H amino acid substitution over SEQ ID NO:636 and SEQ ID NO:715 with a D923K amino acid substitution over SEQ ID NO:636.

Example 27

[1156] Driver constructs encoding engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOS: 790-805 described, hereinbelow were generated and tested with the GFP reporter construct having a nucleic acid sequence set forth in SEQ ID NO: 581.

[1157] SEQ ID NO:790 (encoded by SEQ ID NO: 10790) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, E399M, D555K, D923K, T1011S, I1219N, A1282G. The DNA encoding the R2-TG polypeptide is codon optimized for high frequency codons.

[1158] SEQ ID NO:791 (encoded by SEQ ID NO: 10791) sets forth the amino acid sequence of a R2-TG polypeptide and further includes amino acid substitutions H1240A, Y1311A, R1337A.

[1159] SEQ ID NO:792 (encoded by SEQ ID NO: 10792) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and R866L.

[1160] SEQ ID NO:793 (encoded by SEQ ID NO: 10793) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and F973Q.

[1161] SEQ ID NO:794 (encoded by SEQ ID NO: 10794) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and H544Q.

[1162] SEQ ID NO:795 (encoded by SEQ ID NO: 10795) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and T680S.

[1163] SEQ ID NO:796 (encoded by SEQ ID NO: 10796) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and T603A.

[1164] SEQ ID NO:797 (encoded by SEQ ID NO: 10797) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and Y477H.

[1165] SEQ ID NO:798 (encoded by SEQ ID NO: 10798) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and I1003L.

[1166] SEQ ID NO:799 (encoded by SEQ ID NO: 10799) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and R355K.

[1167] SEQ ID NO:800 (encoded by SEQ ID NO: 10800) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and N301G.

[1168] SEQ ID NO:801 (encoded by SEQ ID NO: 10801) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and K508Q.

[1169] SEQ ID NO:802 (encoded by SEQ ID NO: 10802) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and G217P.

[1170] SEQ ID NO:803 (encoded by SEQ ID NO: 10803) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and E812K.

[1171] SEQ ID NO:804 (encoded by SEQ ID NO: 10804) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and F956I.

[1172] SEQ ID NO:805 (encoded by SEQ ID NO: 10805) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and D1255G.

[1173] SEQ ID NO:715 (encoded by SEQ ID NO: 10715) is described herein, supra.

[1174] FIGS. 27A and 27B are bar graphs showing results of integration assays using engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOs: 790-805. The engineered R2-TG drivers were used with the GFP reporter having nucleic acid sequence set forth in SEQ ID NO:581.

[1175] All of the engineered R2-TG drivers described in this example exhibited improved activity above that of WT R2-TG control (SEQ ID NO:1 +SEQ ID NO:175), except for SEQ ID NO:791. Several modifications showed improved activity over the engineered R2-TG driver and reporter (SEQ ID NO:636 +SEQ ID NO:581), which are used as positive control. Two engineered R2-TG drivers based on SEQ ID NO:636 SEQ ID NO:792 and SEQ ID NO:793 did not exhibit improved activity but were still more active than the WT. The R2-TG drivers with improved activity over the WT and most of them over SEQ ID NO:636 included SEQ ID NO:794, SEQ ID NO:795, SEQ ID NO:796, SEQ ID NO:797, SEQ ID NO:798, SEQ ID NO:799, SEQ ID NO:800, SEQ ID NO:801, SEQ ID NO:802, SEQ ID NO:803, SEQ ID NO:804, and SEQ ID NO:805.

Example 28

[1176] Driver constructs encoding engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOS: 806-821 described, hereinbelow were generated and tested with the GFP reporter construct having a nucleic acid sequence set forth in SEQ ID NO: 581.

[1177] SEQ ID NO:806 (encoded by SEQ ID NO: 10806) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and H1180Q.

[1178] SEQ ID NO:807 (encoded by SEQ ID NO: 10807) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and T966A.

[1179] SEQ ID NO:808 (encoded by SEQ ID NO: 10808) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and L316H.

[1180] SEQ ID NO:809 (encoded by SEQ ID NO: 10809) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and K1151H.

[1181] SEQ ID NO:810 (encoded by SEQ ID NO: 10810) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and E194G.

[1182] SEQ ID NO:811 (encoded by SEQ ID NO: 10811) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and L414G.

[1183] SEQ ID NO:812 (encoded by SEQ ID NO: 10812) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and R574S.

[1184] SEQ ID NO:813 (encoded by SEQ ID NO: 10813) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and F186G.

[1185] SEQ ID NO:814 (encoded by SEQ ID NO: 10814) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and T969E.

[1186] SEQ ID NO:815 (encoded by SEQ ID NO: 10815) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and G432Q.

[1187] SEQ ID NO:816 (encoded by SEQ ID NO: 10816) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and P985R.

[1188] SEQ ID NO:817 (encoded by SEQ ID NO: 10817) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and D591Q.

[1189] SEQ ID NO:818 (encoded by SEQ ID NO: 10818) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and M553R.

[1190] SEQ ID NO:819 (encoded by SEQ ID NO: 10819) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and T174K.

[1191] SEQ ID NO:820 (encoded by SEQ ID NO: 10820) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and D534Y.

[1192] SEQ ID NO:821 (encoded by SEQ ID NO: 10821) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and D591W.

[1193] SEQ ID NO:715 (encoded by SEQ ID NO: 10715) is described hereinabove.

[1194] FIGS. 28A and 28B are bar graphs showing results of integration assays using engineered R2-TG drivers having amino acid sequences in SEQ ID NOs: 806-821. The engineered R2-TG drivers were used with the GFP reporter construct SEQ ID NO:581.

[1195] All the engineered R2-TG drivers described in this example exhibited improved activity above that of WT R2-TG (SEQ ID NO:1 +SEQ ID NO:175), except for SEQ ID NO:816. Several engineered drivers with modifications described herein exhibited improved activity over the engineered R2-TG driver and reporter (SEQ ID NO:636 +SEQ ID NO:581), which are used as another positive control beside the R2-TG WT driver and reporter (SEQ ID NO:1 +SEQ ID NO:175). SEQ ID NO:715 with D932K amino acid substitution was the template for the constructs. Only SEQ ID NO:811 improved the activity by ~20% over the base SEQ ID NO:715 construct, while the others had ~6-8 improved activity over the WT control (SEQ ID NO:1+SEQ ID NO:175).

Example 29

[1196] Driver constructs encoding engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOS: 822-837 described, hereinbelow were generated and tested with the GFP reporter construct having a nucleic acid sequence set forth in SEQ ID NO: 581.

[1197] SEQ ID NO:822 (encoded by SEQ ID NO: 10822) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and D972Q.

[1198] SEQ ID NO:823 (encoded by SEQ ID NO: 10823) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and L975R.

[1199] SEQ ID NO:824 (encoded by SEQ ID NO: 10824) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and C587K.

[1200] SEQ ID NO:825 (encoded by SEQ ID NO: 10825) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and I978K.

[1201] SEQ ID NO:826 (encoded by SEQ ID NO: 10826) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and D591R.

[1202] SEQ ID NO:827 (encoded by SEQ ID NO: 10827) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and D182R.

[1203] SEQ ID NO:828 (encoded by SEQ ID NO: 10828) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and A1024K.

[1204] SEQ ID NO:829 (encoded by SEQ ID NO: 10829) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and P985K.

[1205] SEQ ID NO:830 (encoded by SEQ ID NO: 10830) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and G172R.

[1206] SEQ ID NO:831 (encoded by SEQ ID NO: 10831) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and S178R.

[1207] SEQ ID NO:832 (encoded by SEQ ID NO: 10832) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and G172K.

[1208] SEQ ID NO:833 (encoded by SEQ ID NO: 10833) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and S177W.

[1209] SEQ ID NO:834 (encoded by SEQ ID NO: 10834) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and H544R.

[1210] SEQ ID NO:835 (encoded by SEQ ID NO: 10835) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and G575H.

[1211] SEQ ID NO:836 (encoded by SEQ ID NO: 10836) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and F1135T.

[1212] SEQ ID NO:837 (encoded by SEQ ID NO: 10837) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and K576M.

[1213] SEQ ID NO:715 (encoded by SEQ ID NO: 10715) is described hereinabove.

[1214] FIGS. 29A and 29B are bar graphs showing results of integration assays using engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOs: 822-837. The R2-TG drivers were used with the GFP reporter having nucleic acid sequence set forth in SEQ ID NO:581.

[1215] Most of the engineered R2-TG drivers described in this example exhibited improved activity above that of wild-type R2-TG (SEQ ID NO:1 +SEQ ID NO:175). Several engineered drivers with modifications described herein exhibited improved activity over the positive controls engineered R2-TG driver and reporter (SEQ ID NO:636 +SEQ ID NO:581) and R2-TG WT driver and reporter (SEQ ID NO:1 +SEQ ID NO:175). SEQ ID NO:715 with D932K amino acid substitution served as the base polypeptide for polypeptides in this example. Only SEQ ID NO:834 exhibited improved the activity by ~20% over the base SEQ ID NO:715 construct, while the other improving constructs had ~3-4 fold improved activity over the WT control (SEQ ID NO:1+SEQ ID NO:175).

[1216] SEQ ID NO:825 and SEQ ID NO:829 exhibited decreased activity.

Example 30

[1217] Driver constructs encoding engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOS: 838-852 described, hereinbelow were generated and tested with the GFP reporter construct having a nucleic acid sequence set forth in SEQ ID NO: 581. SEQ ID NO:715 with D932K amino acid substitution served as template for the polypeptides in this example

[1218] SEQ ID NO:838 (encoded by SEQ ID NO: 10838) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and F782W.

[1219] SEQ ID NO:839 (encoded by SEQ ID NO: 10839) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and N1140R.

[1220] SEQ ID NO:840 (encoded by SEQ ID NO: 10840) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and D1133R.

[1221] SEQ ID NO:841 (encoded by SEQ ID NO: 10841) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and D538R.

[1222] SEQ ID NO:842 (encoded by SEQ ID NO: 10842) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and M644R.

[1223] SEQ ID NO:843 (encoded by SEQ ID NO: 10843) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal

fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and A649R.

[1224] SEQ ID NO:844 (encoded by SEQ ID NO: 10844) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and E1125R.

[1225] SEQ ID NO:845 (encoded by SEQ ID NO: 10845) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and P843R.

[1226] SEQ ID NO:846 (encoded by SEQ ID NO: 10846) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and F1135Y.

[1227] SEQ ID NO:847 (encoded by SEQ ID NO: 10847) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and C750K.

[1228] SEQ ID NO:848 (encoded by SEQ ID NO: 10848) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and F1135Y.

[1229] SEQ ID NO:849 (encoded by SEQ ID NO: 10849) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and L1057R.

[1230] SEQ ID NO:850 (encoded by SEQ ID NO: 10850) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal

fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and H1006K.

[1231] SEQ ID NO:851 (encoded by SEQ ID NO: 10851) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and L582K.

[1232] SEQ ID NO:852 (encoded by SEQ ID NO: 10852) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and F876R.

[1233] SEQ ID NO:715 (encoded by SEQ ID NO: 10715) is described hereinabove.

[1234] FIGS. 30A and 30B are bar graphs showing results of integration assays having amino acid sequences set forth in SEQ ID NOs: 838-852.

[1235] Most of the engineered R2-TG drivers described in this example exhibited improved activity above that of WT R2-TG control (SEQ ID NO:1 +SEQ ID NO:175). Several engineered drivers with modifications described herein exhibited improved activity over the positive engineered R2-TG driver and reporter control (SEQ ID NO:636 +SEQ ID NO:581). All the improving constructs had ~3-4 fold improved activity over the WT R2-TG control (SEQ ID NO:1+SEQ ID NO:175).

[1236] SEQ ID NO:838, SEQ ID NO:845, SEQ ID NO:847 and SEQ ID NO: 852 had amino acid substitutions that decreased the activity to below the WT control activity.

Example 31

[1237] Driver constructs encoding engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOS: 853-868 described, hereinbelow were generated and tested with the GFP reporter construct having a nucleic acid sequence set forth in SEQ ID NO: 581. SEQ ID NO:715 with a D932K amino acid substitution served as template for the driver constructs.

[1238] SEQ ID NO:853 (encoded by SEQ ID NO: 10853) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and T728K.

[1239] SEQ ID NO:854 (encoded by SEQ ID NO: 10854) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and P1116K.

[1240] SEQ ID NO:855 (encoded by SEQ ID NO: 10855) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and L243K.

[1241] SEQ ID NO:856 (encoded by SEQ ID NO: 10856) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and E247K.

[1242] SEQ ID NO:857 (encoded by SEQ ID NO: 10857) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and L249F.

[1243] SEQ ID NO:858 (encoded by SEQ ID NO: 10858) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and K255R.

[1244] SEQ ID NO:859 (encoded by SEQ ID NO: 10859) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and V258I.

[1245] SEQ ID NO:860 (encoded by SEQ ID NO: 10860) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and F260Y.

[1246] SEQ ID NO:861 (encoded by SEQ ID NO: 10861) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and Y272H.

[1247] SEQ ID NO:862 (encoded by SEQ ID NO: 10862) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and P291A.

[1248] SEQ ID NO:863 (encoded by SEQ ID NO: 10863) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and D303S.

[1249] SEQ ID NO:864 (encoded by SEQ ID NO: 10864) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and I308R.

[1250] SEQ ID NO:865 (encoded by SEQ ID NO: 10865) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and G311S.

[1251] SEQ ID NO:866 (encoded by SEQ ID NO: 10866) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and I327L.

[1252] SEQ ID NO:867 (encoded by SEQ ID NO: 10867) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and K333R.

[1253] SEQ ID NO:868 (encoded by SEQ ID NO: 10868) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and G339R.

[1254] SEQ ID NO:715 (encoded by SEQ ID NO: 10715) is described hereinabove.

[1255] FIGS. 31A and 31B are bar graphs showing results of integration assays using engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOs: 853-868. The engineered R2-TG drivers were used with the GFP reporter having nucleic acid sequence set forth in SEQ ID NO:581.

[1256] All engineered R2-TG drivers described in this example exhibited improved activity above that of wild-type R2-TG (SEQ ID NO:1 +SEQ ID NO:175). Several engineered drivers with modifications described herein exhibited improved activity over the engineered R2-TG driver and reporter control (SEQ ID NO:636+SEQ ID NO:581). SEQ ID NO:865 and SEQ ID NO:867 improved the activity by ~20% over the base SEQ ID NO:715 construct, while the other improving constructs had ~1.5-2.25 fold improved activity over the WT control (SEQ ID NO:1+SEQ ID NO:175).

Example 32

[1257] Driver construct having an amino acid sequence set forth in SEQ ID NOS: 869-883 described, hereinbelow were generated and tested with the GFP reporter construct having a nucleic acid sequence set forth in SEQ ID NO: 581. SEQ ID NO:715 with a D932K amino acid substitution served as the template for the driver constructs.

[1258] SEQ ID NO:869 (encoded by SEQ ID NO: 10869) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal

fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and K377R.

[1259] SEQ ID NO:870 (encoded by SEQ ID NO: 10870) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and S838K.

[1260] SEQ ID NO:871 (encoded by SEQ ID NO: 10871) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and L474I.

[1261] SEQ ID NO:872 (encoded by SEQ ID NO: 10872) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and T481I.

[1262] SEQ ID NO:873 (encoded by SEQ ID NO: 10873) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and L482I.

[1263] SEQ ID NO:874 (encoded by SEQ ID NO: 10874) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and F498L.

[1264] SEQ ID NO:875 (encoded by SEQ ID NO: 10875) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and M502L.

[1265] SEQ ID NO:876 (encoded by SEQ ID NO: 10876) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal

fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and Q516R.

[1266] SEQ ID NO:877 (encoded by SEQ ID NO: 10877) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and F519I.

[1267] SEQ ID NO:878 (encoded by SEQ ID NO: 10878) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and G520K.

[1268] SEQ ID NO:879 (encoded by SEQ ID NO: 10879) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and C521L.

[1269] SEQ ID NO:880 (encoded by SEQ ID NO: 10880) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and D534K.

[1270] SEQ ID NO:881 (encoded by SEQ ID NO: 10881) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and P545K.

[1271] SEQ ID NO:882 (encoded by SEQ ID NO: 10882) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and A557R.

[1272] SEQ ID NO:883 (encoded by SEQ ID NO: 10883) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal

fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and L564R.

[1273] SEQ ID NO:715 (encoded by SEQ ID NO: 10715) is described hereinabove.

[1274] FIGS. 32A and 32B are bar graphs showing results of integration assays using engineered R2-TG driver having amino acid sequences set forth in SEQ ID NOS: 869-883.

[1275] All engineered R2-TG drivers described in this example exhibited improved activity above that of WT R2-TG control (SEQ ID NO:1 +SEQ ID NO:175), except for SEQ ID NO:882 that had similar activity to the WT R2-TG control. Several engineered drivers with modifications described herein exhibited improved activity over the engineered R2-TG driver and reporter control (SEQ ID NO:636 +SEQ ID NO:581). The constructs with improved activity had ~3-5 fold improved activity over the WT R2-TG control (SEQ ID NO:1 +SEQ ID NO:175).

Example 33

[1276] Driver constructs encoding engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOS: 884-899 described, hereinbelow were generated and tested with the GFP reporter construct having a nucleic acid sequence set forth in SEQ ID NO: 581. SEQ ID NO:715 with a D932K amino acid substitution served as the template for the driver constructs.

[1277] SEQ ID NO:884 (encoded by SEQ ID NO: 10884) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and D573N.

[1278] SEQ ID NO:885 (encoded by SEQ ID NO: 10885) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and L582I.

[1279] SEQ ID NO:886 (encoded by SEQ ID NO: 10886) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and I585K.

[1280] SEQ ID NO:887 (encoded by SEQ ID NO: 10887) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and I597G.

[1281] SEQ ID NO:888 (encoded by SEQ ID NO: 10888) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and F601Y.

[1282] SEQ ID NO:889 (encoded by SEQ ID NO: 10889) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and F628L.

[1283] SEQ ID NO:890 (encoded by SEQ ID NO: 10890) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and L631P.

[1284] SEQ ID NO:891 (encoded by SEQ ID NO: 10891) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and V641L.

[1285] SEQ ID NO:892 (encoded by SEQ ID NO: 10892) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and A649S.

[1286] SEQ ID NO:893 (encoded by SEQ ID NO: 10893) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and P650A.

[1287] SEQ ID NO:894 (encoded by SEQ ID NO: 10894) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and V660I.

[1288] SEQ ID NO:895 (encoded by SEQ ID NO: 10895) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and V661K.

[1289] SEQ ID NO:896 (encoded by SEQ ID NO: 10896) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and E666K.

[1290] SEQ ID NO:897 (encoded by SEQ ID NO: 10897) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and E672R.

[1291] SEQ ID NO:898 (encoded by SEQ ID NO: 10898) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and F674Y.

[1292] SEQ ID NO:899 (encoded by SEQ ID NO: 10899) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and W677I.

[1293] SEQ ID NO:715 (encoded by SEQ ID NO: 10715) is described hereinabove.

[1294] FIGS. 33A and 33B are bar graphs showing results of integration assays using engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOs: 884-899.

[1295] All engineered R2-TG drivers described in this example exhibited improved activity above that of WT R2-TG control (SEQ ID NO:1 +SEQ ID NO:175), except SEQ ID NO:888 and SEQ ID NO:897 that had similar activity to the WT control. Several engineered drivers with modifications described herein exhibited improved activity over the engineered R2-TG driver and reporter control (SEQ ID NO:636 +SEQ ID NO:581). The constructs had ~1.5-3.5 fold improved activity over the WT R2-TG control (SEQ ID NO:1+SEQ ID NO:175).

Example 34

[1296] Driver constructs encoding engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOS: 900-915 described, hereinbelow were generated and tested with the GFP reporter construct having a nucleic acid sequence set forth in SEQ ID NO: 581. SEQ ID NO:715 with a D932K amino acid substitution served as the template for the driver constructs.

[1297] SEQ ID NO:900 (encoded by SEQ ID NO: 10900) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and N738H.

[1298] SEQ ID NO:901 (encoded by SEQ ID NO: 10901) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and L754I.

[1299] SEQ ID NO:902 (encoded by SEQ ID NO: 10902) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and A764S.

[1300] SEQ ID NO:903 (encoded by SEQ ID NO: 10903) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and V771I.

[1301] SEQ ID NO:904 (encoded by SEQ ID NO: 10904) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and V776L.

[1302] SEQ ID NO:905 (encoded by SEQ ID NO: 10905) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and I778F.

[1303] SEQ ID NO:906 (encoded by SEQ ID NO: 10906) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and F782Y.

[1304] SEQ ID NO:907 (encoded by SEQ ID NO: 10907) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and L794M.

[1305] SEQ ID NO:908 (encoded by SEQ ID NO: 10908) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and E798G.

[1306] SEQ ID NO:909 (encoded by SEQ ID NO: 10909) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and Y811L.

[1307] SEQ ID NO:910 (encoded by SEQ ID NO: 10910) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and K821G.

[1308] SEQ ID NO:911 (encoded by SEQ ID NO: 10911) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and K821R.

[1309] SEQ ID NO:912 (encoded by SEQ ID NO: 10912) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and V835I.

[1310] SEQ ID NO:913 (encoded by SEQ ID NO: 10913) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and L844I.

[1311] SEQ ID NO:914 (encoded by SEQ ID NO: 10914) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and C855R.

[1312] SEQ ID NO:915 (encoded by SEQ ID NO: 10915) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and H865K.

[1313] SEQ ID NO:715 (encoded by SEQ ID NO: 10715) is described hereinabove.

[1314] FIGS. 34A and 34B are bar graphs showing results of integration assays using engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOs: 900-915.

[1315] All engineered R2-TG drivers described in this example exhibited improved activity above that of WT R2-TG control (SEQ ID NO:1 +SEQ ID NO:175), except for SEQ ID NO:906 and SEQ ID NO:909 with lower activity than the WT control. SEQ ID NO:914 had similar activity to the WT control. Several engineered drivers with modifications described herein exhibited improved activity over the engineered R2-TG driver and reporter control (SEQ ID

NO:636 +SEQ ID NO:581), The improved constructs had ~2-4 fold improved activity over the WT R2-TG control (SEQ ID NO:1 +SEQ ID NO:175).

Example 35

[1316] Driver constructs encoding engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOS: 916-931 described, hereinbelow were generated and tested with the GFP reporter construct having a nucleic acid sequence set forth in SEQ ID NO: 581. SEQ ID NO:715 with a D932K amino acid substitution served as the template for the driver constructs.

[1317] SEQ ID NO:916 (encoded by SEQ ID NO: 10916) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and D929N.

[1318] SEQ ID NO:917 (encoded by SEQ ID NO: 10917) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and A932P.

[1319] SEQ ID NO:918 (encoded by SEQ ID NO: 10918) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and W933F.

[1320] SEQ ID NO:919 (encoded by SEQ ID NO: 10919) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and M942L.

[1321] SEQ ID NO:920 (encoded by SEQ ID NO: 10920) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and K950R.

[1322] SEQ ID NO:921 (encoded by SEQ ID NO: 10921) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and Y951H.

[1323] SEQ ID NO:922 (encoded by SEQ ID NO: 10922) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and D957N.

[1324] SEQ ID NO:923 (encoded by SEQ ID NO: 10923) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and S965P.

[1325] SEQ ID NO:924 (encoded by SEQ ID NO: 10924) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and T969E.

[1326] SEQ ID NO:925 (encoded by SEQ ID NO: 10925) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and F973K.

[1327] SEQ ID NO:926 (encoded by SEQ ID NO: 10926) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and Q976K.

[1328] SEQ ID NO:927 (encoded by SEQ ID NO: 10927) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and Q980K.

[1329] SEQ ID NO:928 (encoded by SEQ ID NO: 10928) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and A981R.

[1330] SEQ ID NO:929 (encoded by SEQ ID NO: 10929) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and P982K.

[1331] SEQ ID NO:930 (encoded by SEQ ID NO: 10930) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and Y995F.

[1332] SEQ ID NO:931 (encoded by SEQ ID NO: 10931) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and T996R.

[1333] SEQ ID NO:715 (encoded by SEQ ID NO: 10715) is described hereinabove.

[1334] FIGS. 35A and 35B are bar graphs showing results of integration assays using engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOs: 916-931. The engineered R2-TG drivers were used with the GFP reporter having nucleic acid sequence set forth in SEQ ID NO:581.

[1335] All engineered R2-TG drivers described in this example exhibited improved activity above that of WT R2-TG control (SEQ ID NO:1 +SEQ ID NO:175), except SEQ ID NO:931 which exhibited lower activity than the WT control. Several engineered drivers with modifications described herein exhibited improved activity over the engineered R2-TG driver and reporter control (SEQ ID NO:636 +SEQ ID NO:581). SEQ ID NO:918, SEQ ID NO:926 and SEQ ID NO: 931 exhibited ~20% improved activity over SEQ ID NO:715, while the other improving constructs had ~2-6 improved activity over the WT control (SEQ ID NO:1 +SEQ ID NO:175).

Example 36

[1336] Driver constructs encoding engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOS: 932-948 described, hereinbelow were generated and tested with the GFP reporter construct having a nucleic acid sequence set forth in SEQ ID NO: 581. SEQ ID NO:715 with a D932K amino acid substitution served as the template for the driver constructs.

[1337] SEQ ID NO:932 (encoded by SEQ ID NO: 10932) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and L1014I.

[1338] SEQ ID NO:933 (encoded by SEQ ID NO: 10933) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and T1023K.

[1339] SEQ ID NO:934 (encoded by SEQ ID NO: 10934) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and E1027K.

[1340] SEQ ID NO:935 (encoded by SEQ ID NO: 10935) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and D1037N.

[1341] SEQ ID NO:936 (encoded by SEQ ID NO: 10936) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and A1038G.

[1342] SEQ ID NO:937 (encoded by SEQ ID NO: 10937) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal

fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and Y1041H.

[1343] SEQ ID NO:938 (encoded by SEQ ID NO: 10938) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and Y1041H.

[1344] SEQ ID NO:939 (encoded by SEQ ID NO: 10939) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and K1053R.

[1345] SEQ ID NO:940 (encoded by SEQ ID NO: 10940) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and V1061R.

[1346] SEQ ID NO:941 (encoded by SEQ ID NO: 10941) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and K1078R.

[1347] SEQ ID NO:942 (encoded by SEQ ID NO: 10942) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and F1080L.

[1348] SEQ ID NO:943 (encoded by SEQ ID NO: 10943) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and E1082K.

[1349] SEQ ID NO:944 (encoded by SEQ ID NO: 10944) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal

fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and K1092R.

[1350] SEQ ID NO:945 (encoded by SEQ ID NO: 10945) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and G1098R.

[1351] SEQ ID NO:946 (encoded by SEQ ID NO: 10946) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and E1127K.

[1352] SEQ ID NO:947 (encoded by SEQ ID NO: 10947) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and K1132R.

[1353] SEQ ID NO:948 (encoded by SEQ ID NO: 10948) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and Q1155H.

[1354] SEQ ID NO:715 (encoded by SEQ ID NO: 10715) is described hereinabove.

[1355] FIGS. 36A and 36B are bar graphs showing results of integration assays using engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOs: 932-948. The engineered R2-TG drivers were used with the GFP reporter having nucleic acid sequence set forth in SEQ ID NO:581.

[1356] All engineered R2-TG drivers described in this example exhibited improved activity above that of WT R2-TG control (SEQ ID NO:1 +SEQ ID NO:175), except SEQ ID NO:939, SEQ ID NO:941 and SEQ ID NO:945 that had lower activity than the WT control. Several engineered drivers with modifications described herein exhibited improved activity over the engineered R2-TG driver and reporter control (SEQ ID NO:636 +SEQ ID NO:581), SEQ ID

NO:936 exhibited improved the activity by ~20% over the base SEQ ID NO:715 while the other improving constructs had ~1.5-4 improved activity over the WT control (SEQ ID NO:1+SEQ ID NO:175).

Example 37

[1357] Driver constructs encoding engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOS: 949-963 described, hereinbelow were generated and tested with the GFP reporter construct having a nucleic acid sequence set forth in SEQ ID NO: 581. SEQ ID NO:715 with a D932K amino acid substitution served as the template for the driver constructs.

[1358] SEQ ID NO:949 (encoded by SEQ ID NO: 10949) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and Q1188K.

[1359] SEQ ID NO:950 (encoded by SEQ ID NO: 10950) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and Q1233K.

[1360] SEQ ID NO:951 (encoded by SEQ ID NO: 10951) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and E1249K.

[1361] SEQ ID NO:952 (encoded by SEQ ID NO: 10952) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and V1257T.

[1362] SEQ ID NO:953 (encoded by SEQ ID NO: 10953) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and V1287I.

[1363] SEQ ID NO:954 (encoded by SEQ ID NO: 10954) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and V1291I.

[1364] SEQ ID NO:955 (encoded by SEQ ID NO: 10955) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and H1313P.

[1365] SEQ ID NO:956 (encoded by SEQ ID NO: 10956) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and V1318I.

[1366] SEQ ID NO:957 (encoded by SEQ ID NO: 10957) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and K1355R.

[1367] SEQ ID NO:958 (encoded by SEQ ID NO: 10958) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and E1363K.

[1368] SEQ ID NO:959 (encoded by SEQ ID NO: 10959) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and T1367R.

[1369] SEQ ID NO:960 (encoded by SEQ ID NO: 10960) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and H1378R.

[1370] SEQ ID NO:961 (encoded by SEQ ID NO: 10961) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and F1380Y.

[1371] SEQ ID NO:962 (encoded by SEQ ID NO: 10962) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and D1075E.

[1372] SEQ ID NO:963 (encoded by SEQ ID NO: 10963) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and T670L.

[1373] SEQ ID NO:715 (encoded by SEQ ID NO: 10715) is described hereinabove.

[1374] FIGS. 37A and 37B are bar graphs showing results of integration assays using engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOs: 949-963. The engineered R2-TG drivers were used with the GFP reporter having nucleic acid sequence set forth in SEQ ID NO:581.

[1375] All engineered R2-TG drivers described in this example exhibited improved activity above that of WT R2-TG control (SEQ ID NO:1 + SEQ ID NO:175). Several engineered drivers with modifications described herein exhibited improved activity over the engineered R2-TG driver and reporter control (SEQ ID NO:+SEQ ID NO:581). SEQ ID NO:949, SEQ ID NO:950, SEQ ID NO:953, SEQ ID NO:954, SEQ ID NO:959 and SEQ ID NO:962 exhibited improved activity up to ~20% over SEQ ID NO: 715, while the other improving constructs had ~3-5 improved activity over the WT control (SEQ ID NO:1+SEQ ID NO:175).

Example 38

[1376] Driver constructs encoding engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOS: 884, 964-972 and 974-980 described, hereinbelow were generated and tested with the GFP reporter construct having a nucleic acid sequence set forth in SEQ ID NO:

581. SEQ ID NO:715 with a D932K amino acid substitution served the template for the driver constructs.

[1377] SEQ ID NO:964 (encoded by SEQ ID NO: 10964) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and D972R.

[1378] SEQ ID NO:965 (encoded by SEQ ID NO: 10965) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and W1028Y.

[1379] SEQ ID NO:966 (encoded by SEQ ID NO: 10966) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and E643R.

[1380] SEQ ID NO:967 (encoded by SEQ ID NO: 10967) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and T994W.

[1381] SEQ ID NO:968 (encoded by SEQ ID NO: 10968) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and D878R.

[1382] SEQ ID NO:969 (encoded by SEQ ID NO: 10969) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and A578K.

[1383] SEQ ID NO:970 (encoded by SEQ ID NO: 10970) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal

fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and I370L.

[1384] SEQ ID NO:971 (encoded by SEQ ID NO: 10971) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and I374L.

[1385] SEQ ID NO:972 (encoded by SEQ ID NO: 10972) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and F570Y.

[1386] SEQ ID NO:884 (encoded by SEQ ID NO: 10884) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and D573N.

[1387] SEQ ID NO:974 (encoded by SEQ ID NO: 10974) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and V687L.

[1388] SEQ ID NO:975 (encoded by SEQ ID NO: 10975) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and C690S.

[1389] SEQ ID NO:976 (encoded by SEQ ID NO: 10976) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and I896L.

[1390] SEQ ID NO:977 (encoded by SEQ ID NO: 10977) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal

fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and C901L.

[1391] SEQ ID NO:978 (encoded by SEQ ID NO: 10978) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and P998S.

[1392] SEQ ID NO:979 (encoded by SEQ ID NO: 10979) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and Y1002H.

[1393] SEQ ID NO:980 (encoded by SEQ ID NO: 10980) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and W1171F.

[1394] SEQ ID NO:715 (encoded by SEQ ID NO: 10715) is described hereinabove.

[1395] FIGS. 38A and 38B are bar graphs showing results of integration assays using engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOs: 884, 964-972 and 974-980. The engineered R2-TG drivers were used with the GFP reporter having nucleic acid sequence set forth in SEQ ID NO:581.

[1396] All engineered R2-TG drivers described in this example exhibited improved activity above that of WT R2-TG control (SEQ ID NO:1 +SEQ ID NO:175), except for SEQ ID NO:968 with reduced activity compared to the WT control. SEQ ID NO:969 and SEQ ID NO:979 exhibited similar activity to the WT control. Several engineered drivers with modifications described herein exhibited improved activity over the engineered R2-TG driver and reporter control (SEQ ID NO:636 +SEQ ID NO:581. SEQ ID NO:967, SEQ ID NO:971, SEQ ID NO:976 and SEQ ID NO:977 exhibited improved activity up to ~20% over SEQ ID NO:715, while the other improving constructs had ~3-5 improved activity over the WT control (SEQ ID NO:1+SEQ ID NO:175).

Example 39

[1397] Driver constructs encoding engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOS: 981-996 described, hereinbelow were generated and tested with the GFP reporter construct having a nucleic acid sequence set forth in SEQ ID NO: 581. SEQ ID NO:715 with a D932K amino acid substitution served as the template for the driver constructs.

[1398] SEQ ID NO:981 (encoded by SEQ ID NO: 1098) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and L1089M.

[1399] SEQ ID NO:982 (encoded by SEQ ID NO: 10982) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and N1219L.

[1400] SEQ ID NO:983 (encoded by SEQ ID NO: 10983) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and E472G.

[1401] SEQ ID NO:984 (encoded by SEQ ID NO: 10984) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and K480R.

[1402] SEQ ID NO:985 (encoded by SEQ ID NO: 10985) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and N562D.

[1403] SEQ ID NO:986 (encoded by SEQ ID NO: 10986) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal

fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and H1320E.

[1404] SEQ ID NO:987 (encoded by SEQ ID NO: 10987) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and K907R.

[1405] SEQ ID NO:988 (encoded by SEQ ID NO: 10988) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and I778F.

[1406] SEQ ID NO:989 (encoded by SEQ ID NO: 10989) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and W1028K.

[1407] SEQ ID NO:990 (encoded by SEQ ID NO: 10990) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and P169R.

[1408] SEQ ID NO:991 (encoded by SEQ ID NO: 10991) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and D927M.

[1409] SEQ ID NO:992 (encoded by SEQ ID NO: 10992) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and C170K.

[1410] SEQ ID NO:993 (encoded by SEQ ID NO: 10993) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal

fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and S177R.

[1411] SEQ ID NO:994 (encoded by SEQ ID NO: 10994) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and A166R.

[1412] SEQ ID NO:995 (encoded by SEQ ID NO: 10995) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and P548R.

[1413] SEQ ID NO:996 (encoded by SEQ ID NO: 10996) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and S204R.

[1414] SEQ ID NO:715 (encoded by SEQ ID NO: 10715) is described hereinabove.

[1415] SEQ ID NO:636 (encoded by SEQ ID NO: 10636) is described hereinabove.

[1416] FIGS. 39A and 39B are bar graphs showing results of integration assays using engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOs: 981-996. The engineered R2-TG drivers were used with the GFP reporter having nucleic acid sequence set forth in SEQ ID NO:581.

[1417] All modified constructs exhibited improved activity above that of WT R2-TG control (SEQ ID NO: 1+SEQ ID NO:175), except for SEQ ID NO:989, which exhibited reduced activity compared to the WT control. Several engineered drivers with modifications described herein exhibited improved activity over the engineered R2-TG driver and reporter control (SEQ ID NO:636 +SEQ ID NO:581). SEQ ID NO:982 and SEQ ID NO:993 exhibited improved activity up to ~20% over SEQ ID NO:715, while the other improving constructs had ~1.5-7 improved activity over the WT control (SEQ ID NO:1 +SEQ ID NO:175).

Example 40

[1418] Driver constructs encoding engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOS: 997-1012 described, hereinbelow were generated and tested with the GFP reporter construct having a nucleic acid sequence set forth in SEQ ID NO: 581. SEQ ID NO:715 with a D932K amino acid substitution served as the template for the driver constructs.

[1419] SEQ ID NO:997 (encoded by SEQ ID NO: 10997) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and D879W.

[1420] SEQ ID NO:998 (encoded by SEQ ID NO: 10998) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and G575R.

[1421] SEQ ID NO:999 (encoded by SEQ ID NO: 10999) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and A1128R.

[1422] SEQ ID NO: 1000 (encoded by SEQ ID NO: 11000) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and D839R.

[1423] SEQ ID NO: 1001 (encoded by SEQ ID NO: 11001) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and E286K.

[1424] SEQ ID NO: 1002 (encoded by SEQ ID NO: 11002) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal

fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and T532K.

[1425] SEQ ID NO: 1003 (encoded by SEQ ID NO: 11003) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and L569R.

[1426] SEQ ID NO: 1004 (encoded by SEQ ID NO: 11004) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and S589R.

[1427] SEQ ID NO: 1005 (encoded by SEQ ID NO: 11005) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and F611Y.

[1428] SEQ ID NO: 1006 (encoded by SEQ ID NO: 11006) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and Q741R.

[1429] SEQ ID NO: 1007 (encoded by SEQ ID NO: 11007) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and H768R.

[1430] SEQ ID NO: 1008 (encoded by SEQ ID NO: 11008) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and A1063L.

[1431] SEQ ID NO: 1009 (encoded by SEQ ID NO: 11009) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal

fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and R1064K.

[1432] SEQ ID NO: 1010 (encoded by SEQ ID NO: 11010) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and D1075E.

[1433] SEQ ID NO: 1011 (encoded by SEQ ID NO: 11011) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and E1127Q.

[1434] SEQ ID NO: 1012 (encoded by SEQ ID NO: 11012) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and T1130K.

[1435] SEQ ID NO:715 (encoded by SEQ ID NO: 10715) is described hereinabove.

[1436] FIGS. 40A and 40B are bar graphs showing results of integration assays using engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOs: 997-1012.

[1437] All engineered R2-TG drivers described in this example exhibited improved activity above that of WT R2-TG control (SEQ ID NO: 1+SEQ ID NO:175), except for SEQ ID NO:997, SEQ ID NO: 1000, SEQ ID NO: 1006 and SEQ ID NO: 1007 which have reduced activity compared to the WT control. Several engineered drivers described herein exhibited improved activity over the engineered R2-TG driver and reporter control (SEQ ID NO:636 +SEQ ID NO:581). SEQ ID NO: 1004, SEQ ID NO: 1005, SEQ ID NO: 1010 and SEQ ID NO: 1012 exhibited improved activity up to ~30% over SEQ ID NO:715, while the other improving constructs had ~1.5-4.5 improved activity over the WT control (SEQ ID NO: 1 + SEQ ID NO:175).

Example 41

[1438] Driver constructs encoding engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOS: 1013-1026 described, hereinbelow were generated and tested with the GFP reporter construct having a nucleic acid sequence set forth in SEQ ID NO: 581.

[1439] SEQ ID NO: 1013 (encoded by SEQ ID NO: 11013) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, E399M, D555K, D923K, T1011S, I1219N, A1282G and T989L. The DNA encoding the R2-TG polypeptide is codon optimized for high frequency codons.

[1440] SEQ ID NO: 1014 (encoded by SEQ ID NO: 11014) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, HCV polypeptide, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, E399M, D555K, D923K, T1011S, I1219N, A1282G and T989L. The DNA encoding the R2-TG polypeptide is codon optimized for high frequency codons.

[1441] SEQ ID NO: 1015 (encoded by SEQ ID NO: 11015) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, EIF4a polypeptide, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, E399M, D555K, D923K, T1011S, I1219N, A1282G and T989L. The DNA encoding the R2-TG polypeptide is codon optimized for high frequency codons.

[1442] SEQ ID NO: 1016 (encoded by SEQ ID NO: 11016) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, RecT polypeptide, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, E399M, D555K, D923K, T1011S, I1219N, A1282G and T989L. The DNA encoding the R2-TG polypeptide is codon optimized for high frequency codons.

[1443] SEQ ID NO: 1017 (encoded by SEQ ID NO: 11017) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal

fusion of a SV40 NLS, CTIP polypeptide, RecT polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, E399M, D555K, D923K, T1011S, I1219N, A1282G and T989L. The DNA encoding the R2-TG polypeptide is codon optimized for high frequency codons.

[1444] SEQ ID NO: 1018 (encoded by SEQ ID NO: 11018) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, RecT polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, E399M, D555K, D923K, T1011S, I1219N, A1282G, T989L and N269R. The DNA encoding the R2-TG polypeptide is codon optimized for high frequency codons.

[1445] SEQ ID NO: 1019 (encoded by SEQ ID NO: 11019) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, RecT polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, E399M, D555K, D923K, T1011S, I1219N, A1282G, T989L and D1133P. The DNA encoding the engineered R2-TG polypeptide is codon optimized for high frequency codons.

[1446] SEQ ID NO:1020 (encoded by SEQ ID NO: 11020) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, RecT polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, E399M, D555K, D923K, T1011S, I1219N, A1282G, T989L and K1085R. The DNA encoding the R2-TG polypeptide is codon optimized for high frequency codons.

[1447] SEQ ID NO:1021 (encoded by SEQ ID NO: 11021) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, RecT polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, E399M, D555K, D923K, T1011S, I1219N, A1282G, T989L and R1177G. The DNA encoding the R2-TG polypeptide is codon optimized for high frequency codons.

[1448] SEQ ID NO:1022 (encoded by SEQ ID NO: 11022) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, RecT polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, E399M, D555K, D923K, T1011S, I1219N, A1282G, T989L and H423Q. The DNA encoding the R2-TG polypeptide is codon optimized for high frequency codons.

[1449] SEQ ID NO:1023 (encoded by SEQ ID NO: 11023) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue E218) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, E399M, D555K, D923K, T1011S, I1219N, A1282G and T989L. The DNA encoding the R2-TG polypeptide is codon optimized for high frequency codons.

[1450] SEQ ID NO:1024 (encoded by SEQ ID NO: 11024) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, E399M, D555K, D923K, T1011S, I1219N, A1282G and R1337A. The DNA encoding the R2-TG polypeptide is codon optimized for high frequency codons.

[1451] SEQ ID NO:1025 (encoded by SEQ ID NO: 11025) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, HSPB8a peptide, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K.

[1452] SEQ ID NO:1026 (encoded by SEQ ID NO: 11026) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, HSB1L peptide, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K.

[1453] SEQ ID NO: 1012 (encoded by SEQ ID NO: 11012) is described hereinabove.

[1454] SEQ ID NO:915 (encoded by SEQ ID NO: 10915) is described hereinabove.

[1455] SEQ ID NO:715 (encoded by SEQ ID NO: 10715) is described hereinabove.

[1456] FIGS. 41A and 41B are bar graphs showing results of integration assays using engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOs: 1013-1026. The engineered R2-TG drivers were used with the GFP reporter having nucleic acid sequence set forth in SEQ ID NO:581.

[1457] All engineered R2-TG drivers described in this example exhibited improved activity above that of WT R2-TG control (SEQ ID NO:1 +SEQ ID NO:175), except SEQ ID NO:1024 and SEQ ID NO:915 with reduced activity compared to WT control. Several engineered drivers described herein exhibited improved activity over the engineered R2-TG driver and reporter control (SEQ ID NO:636 +SEQ ID NO:581). SEQ ID NO:715 with a D932K amino acid substitution had the highest activity as a reference in this experiment. SEQ ID NO: 1013, SEQ ID NO: 1014, SEQ ID NO: 1015, SEQ ID NO: 1017, SEQ ID NO: 1018, SEQ ID NO: 1019, SEQ ID NO:1088, SEQ ID NO:1021, SEQ ID NO:1022 and SEQ ID NO: 1023 exhibited improved the activity up to ~30-100% over SEQ ID NO:715. Other improving constructs had ~1.5-6 improved activity over the WT control (SEQ ID NO:1+SEQ ID NO:175).

Example 42

[1458] Driver constructs encoding engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOS: 1028-1044 described, hereinbelow were generated and tested with the GFP reporter construct having a nucleic acid sequence set forth in SEQ ID NO: 581.

[1459] SEQ ID NO:1028 (encoded by SEQ ID NO: 11028) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and V727M.

[1460] SEQ ID NO:1029 (encoded by SEQ ID NO: 11029) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and T759G.

[1461] SEQ ID NO:1030 (encoded by SEQ ID NO: 11030) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal

fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and Q795R.

[1462] SEQ ID NO:1031 (encoded by SEQ ID NO: 11031) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and I814C.

[1463] SEQ ID NO:1032 (encoded by SEQ ID NO: 11032) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and Y817R.

[1464] SEQ ID NO:1033 (encoded by SEQ ID NO: 11033) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and T819E.

[1465] SEQ ID NO:1034 (encoded by SEQ ID NO: 11034) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and H825K.

[1466] SEQ ID NO:1035 (encoded by SEQ ID NO: 11035) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and Q868G.

[1467] SEQ ID NO:1036 (encoded by SEQ ID NO: 11036) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and Q868D.

[1468] SEQ ID NO:1037 (encoded by SEQ ID NO: 11037) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal

fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and Q911K.

[1469] SEQ ID NO:1038 (encoded by SEQ ID NO: 11038) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and Q911S.

[1470] SEQ ID NO:1039 (encoded by SEQ ID NO: 11039) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and A931P.

[1471] SEQ ID NO:1040 (encoded by SEQ ID NO: 11040) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and D944G.

[1472] SEQ ID NO:1041 (encoded by SEQ ID NO: 11041) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and F973S.

[1473] SEQ ID NO:1042 (encoded by SEQ ID NO: 11042) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and A1012S.

[1474] SEQ ID NO:1043 (encoded by SEQ ID NO: 11043) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and L657Y.

[1475] SEQ ID NO:1044 (encoded by SEQ ID NO: 11044) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal

fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and S699G.

[1476] FIGS. 42A and 42B are bar graphs showing results of integration assays using engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOs: 1028-1044.

[1477] All of the engineered R2-TG drivers described in this example exhibited improved activity above that of WT R2-TG control (SEQ ID NO:1 +SEQ ID NO:175). Several of the engineered drivers described herein exhibited improved activity over the engineered R2-TG driver and reporter control (SEQ ID NO:636 +SEQ ID NO:581): SEQ ID NO:1028, SEQ ID NO:1030, SEQ ID NO:1031, SEQ ID NO:1033, SEQ ID NO:1034, SEQ ID NO:1035, SEQ ID NO:1036, SEQ ID NO:1039, SEQ ID NO:1040, SEQ ID NO:1041, SEQ ID NO:1042. The drivers having improved activity had ~3-6 fold improvement over the WT driver and reporter control (SEQ ID NO:1 +SEQ ID NO:175) and up to 40% improvement over the engineered R2-TG control (SEQ ID NO:636 +SEQ ID NO:581).

Example 43

[1478] Driver constructs encoding engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOS: 1045-1061 described, hereinbelow were generated and tested with the GFP reporter construct having a nucleic acid sequence set forth in SEQ ID NO: 581.

[1479] SEQ ID NO:1045 (encoded by SEQ ID NO: 11045) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and I707P.

[1480] SEQ ID NO:1046 (encoded by SEQ ID NO: 11046) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and W710Y.

[1481] SEQ ID NO:1047 (encoded by SEQ ID NO: 11047) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal

fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and T714A.

[1482] SEQ ID NO:1048 (encoded by SEQ ID NO: 11048) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and I718T.

[1483] SEQ ID NO:1049 (encoded by SEQ ID NO: 11049) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and I726V.

[1484] SEQ ID NO:1050 (encoded by SEQ ID NO: 11050) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and A779S.

[1485] SEQ ID NO:1051 (encoded by SEQ ID NO: 11051) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and N813G.

[1486] SEQ ID NO:1052 (encoded by SEQ ID NO: 11052) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and V881T.

[1487] SEQ ID NO:1053 (encoded by SEQ ID NO: 11053) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and L906M.

[1488] SEQ ID NO:1054 (encoded by SEQ ID NO: 11054) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal

fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and S648K.

[1489] SEQ ID NO:1055 (encoded by SEQ ID NO: 11055) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and S699K.

[1490] SEQ ID NO:1056 (encoded by SEQ ID NO: 11056) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and S717P.

[1491] SEQ ID NO:1057 (encoded by SEQ ID NO: 11057) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and A748R.

[1492] SEQ ID NO:1058 (encoded by SEQ ID NO: 11058) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and A779K.

[1493] SEQ ID NO:1059 (encoded by SEQ ID NO: 11059) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and D783G.

[1494] SEQ ID NO:1060 (encoded by SEQ ID NO: 11060) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and D851N.

[1495] SEQ ID NO:1061 (encoded by SEQ ID NO: 11061) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal

fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and D800P.

[1496] FIGS. 43A and 43B are bar graphs showing results of integration assays using engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOs: 1045-1061. The engineered R2-TG drivers were used with the GFP reporter having nucleic acid sequence set forth in SEQ ID NO:581.

[1497] All engineered R2-TG drivers described in this example exhibited improved activity above that of WT R2-TG control (SEQ ID NO:1 +SEQ ID NO:175), except for SEQ ID NO:1059, which had similar activity to the WT control (SEQ ID NO:1 + SEQ ID NO:175). Several engineered drivers with modifications described herein exhibited improved activity over the engineered R2-TG driver and reporter control (SEQ ID NO:636+SEQ ID NO:581): SEQ ID NO:1046 SEQ ID NO:1048 SEQ ID NO:1049 SEQ ID NO:1051, SEQ ID NO:1052 SEQ ID NO:1054. The drivers had ~1.5-5.5 fold improvement over the WT driver and reporter (SEQ ID NO:1 +SEQ ID NO:175) and up to 35% improvement over the engineered R2-TG control (SEQ ID NO:636 +SEQ ID NO:581).

Example 44

[1498] Driver constructs encoding engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOS: 1062-1076 described, hereinbelow were generated and tested with the GFP reporter construct having a nucleic acid sequence set forth in SEQ ID NO: 581.

[1499] SEQ ID NO:1062 (encoded by SEQ ID NO: 11062) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and E949Y.

[1500] SEQ ID NO:1063 (encoded by SEQ ID NO: 11063) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and E949Y. CTIP polypeptide includes T15I amino acid substitution.

[1501] SEQ ID NO:1064 (encoded by SEQ ID NO: 11064) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and E949Y. CTIP polypeptide includes R123K amino acid substitution.

[1502] SEQ ID NO:1065 (encoded by SEQ ID NO: 11065) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and E949Y. CTIP polypeptide includes P181L amino acid substitution.

[1503] SEQ ID NO:1066 (encoded by SEQ ID NO: 11066) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and E949Y. CTIP polypeptide includes M202L amino acid substitution.

[1504] SEQ ID NO:1067 (encoded by SEQ ID NO: 11067) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and E949Y. CTIP polypeptide includes M202F amino acid substitution.

[1505] SEQ ID NO:1068 (encoded by SEQ ID NO: 11068) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and E949Y. CTIP polypeptide includes M202T amino acid substitution.

[1506] SEQ ID NO:1069 (encoded by SEQ ID NO: 11069) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid

substitutions P280K, D555K, T1011S, I1219N, A1282G and E949Y. CTIP polypeptide includes A237L amino acid substitution.

[1507] SEQ ID NO:1070 (encoded by SEQ ID NO: 11070) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and E949Y. CTIP polypeptide includes A237I amino acid substitution.

[1508] SEQ ID NO:1071 (encoded by SEQ ID NO: 11071) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and E949Y. CTIP polypeptide includes A237N amino acid substitution.

[1509] SEQ ID NO:1072 (encoded by SEQ ID NO: 11072) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and E949Y. CTIP polypeptide includes A237K amino acid substitution.

[1510] SEQ ID NO:1073 (encoded by SEQ ID NO: 11073) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and E949Y. CTIP polypeptide includes N288D amino acid substitution.

[1511] SEQ ID NO:1074 (encoded by SEQ ID NO: 11074) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and E1317Q.

[1512] SEQ ID NO:1075 (encoded by SEQ ID NO: 11075) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal

fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and M1361L.

[1513] SEQ ID NO:1076 (encoded by SEQ ID NO: 11076) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G and K1386R.

[1514] FIGS. 44A and 44B are bar graphs showing results of integration assays using engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOs: 1062-1076.

[1515] All engineered R2-TG drivers described in this example exhibited improved activity above that of WT R2-TG control (SEQ ID NO:1 +SEQ ID NO:175), except SEQ ID NO:1062 that had lower activity than the WT control. Several engineered drivers with modifications described herein exhibited improved activity over the engineered R2-TG driver and reporter control (SEQ ID NO:636+SEQ ID NO:581). The drivers having improved activity had ~1.5-4.5 fold improvement over the WT driver and reporter (SEQ ID NO:1 +SEQ ID NO:175) and up to 210% improvement over the

Example 45

[1516] Driver constructs encoding engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOS: 1077- 1092 described, hereinbelow were generated and tested with the GFP reporter construct having a nucleic acid sequence set forth in SEQ ID NO: 581.

[1517] SEQ ID NO:1077 (encoded by SEQ ID NO: 11077) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, E399M, D555K, D923K, T1011S, I1219N, A1282G, T989L and Y571H. The DNA encoding the R2-TG polypeptide is codon optimized for high frequency codons.

[1518] SEQ ID NO:1078 (encoded by SEQ ID NO: 11078) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, UL42 polypeptide, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, E399M, D555K, D923K, T1011S, I1219N, A1282G,

T989L and Y571H. The DNA encoding the R2-TG polypeptide is codon optimized for high frequency codons.

[1519] SEQ ID NO:1079 (encoded by SEQ ID NO: 11079) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, E399M, D555K, D923K, T1011S, I1219N, A1282G, T989L and K700D. The DNA encoding the R2-TG polypeptide is codon optimized for high frequency codons.

[1520] SEQ ID NO:1080 (encoded by SEQ ID NO: 11080) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, E399M, D555K, D923K, T1011S, I1219N, A1282G, T989L and E214R. The DNA encoding the R2-TG polypeptide is codon optimized for high frequency codons.

[1521] SEQ ID NO:1081(encoded by SEQ ID NO: 11081) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, E399M, D555K, D923K, T1011S, I1219N, A1282G, T989L and D537Y. The DNA encoding the R2-TG polypeptide is codon optimized for high frequency codons.

[1522] SEQ ID NO:1082 (encoded by SEQ ID NO: 11082) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, E399M, D555K, D923K, T1011S, I1219N, A1282G, T989L and S895K. The DNA encoding the R2-TG polypeptide is codon optimized for high frequency codons.

[1523] SEQ ID NO:1083 (encoded by SEQ ID NO: 11083) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, E399M, D555K, D923K, T1011S, I1219N, A1282G, T989L and L1013R. The DNA encoding the R2-TG polypeptide is codon optimized for high frequency codons.

[1524] SEQ ID NO:1084 (encoded by SEQ ID NO: 11084) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal

fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, E399M, D555K, D923K, T1011S, I1219N, A1282G, T989L and E214R, K700D. The DNA encoding the R2-TG polypeptide is codon optimized for high frequency codons.

[1525] SEQ ID NO:1085 (encoded by SEQ ID NO: 11085) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, E399M, D555K, D923K, T1011S, I1219N, A1282G, T989L and E214R, N269R, K700D. The DNA encoding the R2-TG polypeptide is codon optimized for high frequency codons.

[1526] SEQ ID NO:1086 (encoded by SEQ ID NO: 11086) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, E399M, D555K, D923K, T1011S, I1219N, A1282G, T989L and R669I. The DNA encoding the R2-TG polypeptide is codon optimized for high frequency codons.

[1527] SEQ ID NO:1087 (encoded by SEQ ID NO: 11087) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, E399M, D555K, D923K, T1011S, I1219N, A1282G, T989L and F973E. The DNA encoding the R2-TG polypeptide is codon optimized for high frequency codons.

[1528] SEQ ID NO:1088 (encoded by SEQ ID NO: 11088) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, E399M, D555K, D923K, T1011S, I1219N, A1282G, T989L and R669L, D1133P. The DNA encoding the R2-TG polypeptide is codon optimized for high frequency codons.

[1529] SEQ ID NO:1089 (encoded by SEQ ID NO: 11089) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid

substitutions P280K, E399M, D555K, D923K, T1011S, I1219N, A1282G, T989L and D537Y, R669L, D1133P. The DNA encoding the R2-TG polypeptide is codon optimized for high frequency codons.

[1530] SEQ ID NO:1090 (encoded by SEQ ID NO: 11090) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, E399M, D555K, D923K, T1011S, I1219N, A1282G, T989L and K547R. The DNA encoding the R2-TG polypeptide is codon optimized for high frequency codons.

[1531] SEQ ID NO:1091 (encoded by SEQ ID NO: 11091) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, E399M, D555K, D923K, T1011S, I1219N, A1282G, T989L and G257R. The DNA encoding the R2-TG polypeptide is codon optimized for high frequency codons.

[1532] SEQ ID NO:1092 (encoded by SEQ ID NO: 11092) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, E399M, D555K, D923K, T1011S, I1219N, A1282G, T989L and V539G. The DNA encoding the R2-TG polypeptide is codon optimized for high frequency codons.

[1533] SEQ ID NO:715 (encoded by SEQ ID NO: 10715) is described hereinabove.

[1534] FIGS. 45A and 45B are bar graphs showing results of integration assays using engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOs: 1077-1092.

[1535] All engineered R2-TG drivers described in this example exhibited improved activity above that of WT R2-TG control (SEQ ID NO:1 +SEQ ID NO:175). Several engineered drivers with modifications described herein exhibited improved activity over the engineered R2-TG driver and reporter control (SEQ ID NO:636 +SEQ ID NO:581). The drivers having improved activity had ~1.5-3 fold improvement over the engineered R2-TG control (SEQ ID NO:636 +SEQ ID NO:581).

Example 46

[1536] Driver constructs encoding engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOS: 1093-1107 described, hereinbelow were generated and tested with the GFP reporter construct having a nucleic acid sequence set forth in SEQ ID NO: 581.

[1537] SEQ ID NO:1093 (encoded by SEQ ID NO: 11093) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, E399M, D555K, D923K, T1011S, I1219N, A1282G, T989L and N363H. The DNA encoding the R2-TG polypeptide is codon optimized for high frequency codons.

[1538] SEQ ID NO:1094 (encoded by SEQ ID NO: 11094) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, E399M, D555K, D923K, T1011S, I1219N, A1282G, T989L and A442P. The DNA encoding the R2-TG polypeptide is codon optimized for high frequency codons.

[1539] SEQ ID NO:1095 (encoded by SEQ ID NO: 11095) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, E399M, D555K, D923K, T1011S, I1219N, A1282G, T989L and K547R, K1085R. The DNA encoding the R2-TG polypeptide is codon optimized for high frequency codons.

[1540] SEQ ID NO:1096 (encoded by SEQ ID NO: 11096) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, E399M, D555K, D923K, T1011S, I1219N, A1282G, T989L and Y1311A. The DNA encoding the R2-TG polypeptide is codon optimized for high frequency codons.

[1541] SEQ ID NO:1097 (encoded by SEQ ID NO: 11097) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, E399M, D555K, D923K, T1011S, I1219N, A1282G, T989L and Y1311A,

R1337A. The DNA encoding the R2-TG polypeptide is codon optimized for high frequency codons.

[1542] SEQ ID NO:1098 (encoded by SEQ ID NO: 11098) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, E399M, D555K, D923K, T1011S, I1219N, A1282G, T989L and H1240A, Y1311A, R1337A. The DNA encoding the R2-TG polypeptide is codon optimized for high frequency codons.

[1543] SEQ ID NO:1099 (encoded by SEQ ID NO: 11099) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, E399M, D555K, D923K, T1011S, I1219N, A1282G, T989L and H1240A, Y1311A. The DNA encoding the R2-TG polypeptide is codon optimized for high frequency codons.

[1544] SEQ ID NO:1100 (encoded by SEQ ID NO: 11100) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K and N891K.

[1545] SEQ ID NO:1101 (encoded by SEQ ID NO: 11101) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide. R2-TG polypeptide includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K and K1269G.

[1546] SEQ ID NO:1102 (encoded by SEQ ID NO: 11102) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K and R574G.

[1547] SEQ ID NO:1103 (encoded by SEQ ID NO: 11103) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal

fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide. R2-TG polypeptide includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K and D1288A.

[1548] SEQ ID NO:1104 (encoded by SEQ ID NO: 11104) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K and D573R.

[1549] SEQ ID NO:1105 (encoded by SEQ ID NO: 11105) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide. R2-TG polypeptide includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K and T994R.

[1550] SEQ ID NO:1107 (encoded by SEQ ID NO: 11107) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K and V785I.

[1551] SEQ ID NO:924 (encoded by SEQ ID NO: 10924) is described hereinabove.

[1552] SEQ ID NO:715 (encoded by SEQ ID NO: 10715) is described hereinabove.

[1553] FIGS. 46A and 46B are bar graphs showing results of integration assays using engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOs: 1093-1107. The engineered R2-TG drivers were used with the GFP reporter having nucleic acid sequence set forth in SEQ ID NO:581.

[1554] Most of the engineered R2-TG drivers described in this example exhibited improved activity above that of WT R2-TG control (SEQ ID NO:1 +SEQ ID NO:175). Several of the engineered drivers with modifications described herein exhibited improved activity over the engineered R2-TG driver and reporter control (SEQ ID NO:636 +SEQ ID NO:581. The drivers having improved activity are: SEQ ID NO: 1093, SEQ ID NO:1094, SEQ ID NO:1095, SEQ ID NO:1100, SEQ ID NO:1101, SEQ ID NO:1104, SEQ ID NO: 1105, SEQ ID NO:1106, SEQ ID NO:1107 and SEQ ID NO:1108 each exhibited a ~1.5-5 fold improvement over the WT R2-TG control (SEQ ID NO:1 + SEQ ID NO:175).

[1555] Drivers having amino acid sequences set forth in SEQ ID NO:1096, SEQ ID NO:1097, SEQ ID NO:1315, SEQ ID NO:1099 and SEQ ID NO:1103 exhibited reduced activity below WT control levels, seemingly due to the specific amino acid substitutions.

Example 47

[1556] Driver constructs encoding engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOS: 1109-1124 described, hereinbelow were generated and tested with the GFP reporter construct having a nucleic acid sequence set forth in SEQ ID NO: 581.

[1557] SEQ ID NO:1109 (encoded by SEQ ID NO: 11109) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and L175S.

[1558] SEQ ID NO:1110 (encoded by SEQ ID NO: 11110) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and D199S.

[1559] SEQ ID NO:1111 (encoded by SEQ ID NO: 11111) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and V239M.

[1560] SEQ ID NO:1112 (encoded by SEQ ID NO: 11112) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and V299E.

[1561] SEQ ID NO:1113 (encoded by SEQ ID NO: 11113) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and E334A.

[1562] SEQ ID NO:1114 (encoded by SEQ ID NO: 11114) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and G520A.

[1563] SEQ ID NO:1115 (encoded by SEQ ID NO: 11115) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and T532G.

[1564] SEQ ID NO:1116 (encoded by SEQ ID NO: 11116) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and D591G.

[1565] SEQ ID NO:1117 (encoded by SEQ ID NO: 11117) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and S599D.

[1566] SEQ ID NO:1118 (encoded by SEQ ID NO: 11118) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and K622D.

[1567] ID NO:1119 (encoded by SEQ ID NO: 11119) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and T1122L.

[1568] SEQ ID NO:1120 (encoded by SEQ ID NO: 11120) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and L1285I.

[1569] SEQ ID NO:1121 (encoded by SEQ ID NO: 11121) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and I1003L.

[1570] SEQ ID NO:1122 (encoded by SEQ ID NO: 11122) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and L720Y.

[1571] SEQ ID NO:1123 (encoded by SEQ ID NO: 11123) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and K780N.

[1572] SEQ ID NO:1124 (encoded by SEQ ID NO: 11124) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G. CTPI polypeptide includes R32K amino acid substitution.

[1573] SEQ ID NO:643 (encoded by SEQ ID NO: 10643) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and F876Y.

[1574] FIGS. 47A and 47B are bar graphs showing results of integration assays using engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOs: 1109-1124. The engineered R2-TG drivers were used with the GFP reporter having nucleic acid sequence set forth in SEQ ID NO:581.

[1575] All of the engineered R2-TG drivers described in this example exhibited improved activity above that of WT R2-TG control (SEQ ID NO:1 +SEQ ID NO:175) except for SEQ ID NO:1122 that had activity below WT control and SEQ ID NO:1123 that had similar activity to the WT control. Several engineered drivers with modifications described herein exhibited

improved activity over the engineered R2-TG driver and reporter control (SEQ ID NO:636 +SEQ ID NO:581).

[1576] The drivers having improved activity exhibited ~2-4 fold improvement over the WT R2-TG control (SEQ ID NO:1 +SEQ ID NO:175) and up to 30% improvement over the engineered R2-TG (SEQ ID NO:636 + SEQ ID NO:581).

Example 48

[1577] Driver constructs encoding engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOS: 1125-1139 described, hereinbelow were generated and tested with the GFP reporter construct having a nucleic acid sequence set forth in SEQ ID NO: 581.

[1578] SEQ ID NO:1125 (encoded by SEQ ID NO: 11125) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, E399M, D555K, D923K, T1011S, I1219N, A1282G, T989L and E214R, N269R, D537Y, K700D. The DNA encoding the R2-TG polypeptide is codon optimized for high frequency codons.

[1579] SEQ ID NO:1126 (encoded by SEQ ID NO: 11126) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, E399M, D555K, D923K, T1011S, I1219N, A1282G, T989L and D537Y, R669L, F973E and D1133P. The DNA encoding the R2-TG polypeptide is codon optimized for high frequency codons.

[1580] SEQ ID NO:1127 (encoded by SEQ ID NO: 11127) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, E399M, D555K, D923K, T1011S, I1219N, A1282G, T989L and N363H, K547R, K1085R. The DNA encoding the R2-TG polypeptide is codon optimized for high frequency codons.

[1581] SEQ ID NO:1128 (encoded by SEQ ID NO: 11128) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, E399M, D555K, D923K, T1011S, I1219N, A1282G, T989L and G257R, N363H, A442P, K547R, K1085R. The DNA encoding the R2-TG polypeptide is codon optimized for high frequency codons.

[1582] SEQ ID NO:1129 (encoded by SEQ ID NO: 11129) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, E399M, D555K, D923K, T1011S, I1219N, A1282G, T989L and H1240A. The DNA encoding the R2-TG polypeptide is codon optimized for high frequency codons.

[1583] SEQ ID NO:1130 (encoded by SEQ ID NO: 11130) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, E399M, D555K, D923K, T1011S, I1219N, A1282G, T989L and H1240A, R1337A. The DNA encoding the R2-TG polypeptide is codon optimized for high frequency codons.

[1584] SEQ ID NO:1131 (encoded by SEQ ID NO: 11130) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, FUSa peptide, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, D923K, T1011S, I1219N, A1282G, T989L.

[1585] SEQ ID NO:1132 (encoded by SEQ ID NO: 11132) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K and L572K.

[1586] SEQ ID NO:1133 (encoded by SEQ ID NO: 11133) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K and Q1062R.

[1587] SEQ ID NO:1134 (encoded by SEQ ID NO: 11134) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K and L1184I.

[1588] SEQ ID NO:1135 (encoded by SEQ ID NO: 11135) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K and N1228Q.

[1589] SEQ ID NO:1136 (encoded by SEQ ID NO: 11136) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K and F1345N.

[1590] SEQ ID NO:645 (encoded by SEQ ID NO: 10645) is described hereinabove.

[1591] SEQ ID NO:633 (encoded by SEQ ID NO: 10633) is described hereinabove.

[1592] SEQ ID NO:623 (encoded by SEQ ID NO: 10623) is described hereinabove.

[1593] SEQ ID NO:715 (encoded by SEQ ID NO: 10715) is described hereinabove.

[1594] SEQ ID NO:636 (encoded by SEQ ID NO: 10636) is described hereinabove.

[1595] FIGS. 48A and 48B are bar graphs showing results of integration assays using engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOs: 1125-1136. The engineered R2-TG drivers were used with the GFP reporter having nucleic acid sequence set forth in SEQ ID NO:581.

[1596] All the engineered R2-TG drivers described in this example exhibited improved activity above that of WT R2-TG control (SEQ ID NO:1+SEQ ID NO:175), except or SEQ ID NO:1129 and SEQ ID NO:1130 that exhibited activity below WT control. Several engineered drivers with modifications described herein exhibited improved activity over the engineered R2-TG driver and reporter control (SEQ ID NO:636 +SEQ ID NO:581).

[1597] SEQ ID NO:1125, SEQ ID NO:1126, SEQ ID NO:1127, SEQ ID NO:1128, SEQ ID NO:1131, SEQ ID NO:1132, SEQ ID NO:633, SEQ ID NO:623 exhibited improved activity. The drivers having improved activity had 2 fold improvement over the engineered R2-TG control (SEQ ID NO:636 +SEQ ID NO:581) and up to 8 fold improvement over the WT R2-TG control (SEQ ID NO:1 +SEQ ID NO:175).

Example 49

[1598] Driver constructs encoding engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOS: 1137-1148 described, hereinbelow were generated and tested with the GFP reporter construct having a nucleic acid sequence set forth in SEQ ID NO: 175.

[1599] SEQ ID NO:1137 (encoded by SEQ ID NO: 11137) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G. CTIP polypeptide includes E188D amino acid substitution.

[1600] SEQ ID NO:1138 (encoded by SEQ ID NO: 11138) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G. CTIP polypeptide includes E188A amino acid substitution.

[1601] SEQ ID NO:1139 (encoded by SEQ ID NO: 11139) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and R435K.

[1602] SEQ ID NO:1976 (encoded by SEQ ID NO: 11976) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and R435K, L438A.

[1603] SEQ ID NO:1140 (encoded by SEQ ID NO: 11140) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and L438A.

[1604] SEQ ID NO:1141 (encoded by SEQ ID NO: 11141) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and A1191L.

[1605] SEQ ID NO:1977 (encoded by SEQ ID NO: 11977) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and A1191L, P1195G.

[1606] SEQ ID NO:1978 (encoded by SEQ ID NO: 11978) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and P1195G.

[1607] SEQ ID NO:1142 (encoded by SEQ ID NO: 11142) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G. HMGN1 polypeptide includes amino acid substitution D100E.

[1608] SEQ ID NO:1143 (encoded by SEQ ID NO: 11143) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G. HMGN1 polypeptide includes amino acid substitutions D100G.

[1609] SEQ ID NO:1144 (encoded by SEQ ID NO: 11144) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal

fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and R1197K.

[1610] SEQ ID NO:1145 (encoded by SEQ ID NO: 11145) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and R1197K, L1200I.

[1611] SEQ ID NO:1146 (encoded by SEQ ID NO: 11146) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and G1331A.

[1612] SEQ ID NO:1147 (encoded by SEQ ID NO: 11147) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and G1335A.

[1613] SEQ ID NO:1148 (encoded by SEQ ID NO: 11148) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and G1331A, G1335A.

[1614] FIGS. 49A and 49B are bar graphs showing results of integration assays using engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOs: 1137-1148. The engineered R2-TG drivers were used with the GFP reporter construct SEQ ID NO:175). The driver was added to the reporter after 6h or 24h. The drivers were engineered R2-TG protein constructs described above.

[1615] Most of the engineered R2-TG drivers described in this example exhibited improved activity above that of WT R2-TG control (SEQ ID NO:+SEQ ID NO:175) except for SEQ ID NO:1141, SEQ ID NO:1147 and SEQ ID NO:1148, which had activity below WT control and SEQ ID NO:1145 that had a similar activity to the WT control. The template for the amino acid substitutions in this experiment was SEQ ID NO:636 and SEQ ID NO:636 + NO:175 served as a control for activity (integration). SEQ ID NO:1143 had ~2 fold-improved activity over WT

control and ~15% improvement over the engineered driver and reporter control (SEQ ID NO:636 +SEQ ID NO:175). The improvement might be due to extended expression of the protein as a result of lower degradation rates.

Example 50

[1616] Driver constructs encoding engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOS: 1149-1159 and 161-162 described, hereinbelow were generated and tested with the GFP reporter construct having a nucleic acid sequence set forth in SEQ ID NO: 175.

[1617] SEQ ID NO:1149 (encoded by SEQ ID NO: 11149) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and K856A, S860A.

[1618] SEQ ID NO:1150 (encoded by SEQ ID NO: 11150) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and N819S, T892V, S895V.

[1619] SEQ ID NO:1151 (encoded by SEQ ID NO: 11151) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and N819A, T892A, S895A.

[1620] SEQ ID NO:1152 (encoded by SEQ ID NO: 11152) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and E898S, T899V, N902S.

[1621] SEQ ID NO:1153 (encoded by SEQ ID NO: 11153) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and E898A, T899A, N902A.

[1622] SEQ ID NO:1154 (encoded by SEQ ID NO: 11154) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and T934V, G937P.

[1623] SEQ ID NO:1155 (encoded by SEQ ID NO: 11155) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and T934A, G937A.

[1624] SEQ ID NO:1156 (encoded by SEQ ID NO: 11156) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and K1252D, D1255K, W1256G.

[1625] SEQ ID NO:1157 (encoded by SEQ ID NO: 11157) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and K1252A, D1255A, W1256A.

[1626] SEQ ID NO:1158 (encoded by SEQ ID NO: 11158) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and K612D, D616K.

[1627] SEQ ID NO:1159 (encoded by SEQ ID NO: 11159) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and K612A, D616A.

[1628] SEQ ID NO:1161 (encoded by SEQ ID NO: 11161) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and P593G, S595V, E596S, S599V.

[1629] SEQ ID NO:1162 (encoded by SEQ ID NO: 11162) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and P593A, S595A, E596A, S599A.

[1630] FIGS. 50A and 50B are bar graphs showing results of integration assays using engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOs: 1149-1159 and 161-162. The engineered R2-TG drivers were used with the GFP reporter construct SEQ ID NO:175. The driver was added to the reporter after 6h or 24h.

[1631] Most of the engineered R2-TG drivers described in this example improved had a similar activity to the WT control wild-type R2-TG control (SEQ ID NO:1 +SEQ ID NO:175), except SEQ ID NO:1149 and SEQ ID NO:1151, which had activity below WT control level. The base construct for the amino acid substitutions in this experiment was SEQ ID NO:636 and SEQ ID NO:636 + SEQ ID NO:175 were used as another control. SEQ ID NO: 1152 and SEQ ID NO:1153 exhibited ~20-35% improvement over the reference control (SEQ ID NO: 636+SEQ ID NO:175). The improvement might be due to extended expression of the protein because of lower degradation rates.

Example 51

[1632] Driver constructs encoding engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOS: 1163-1170 and 1977-1978 described, hereinbelow were generated and tested with the GFP reporter construct having a nucleic acid sequence set forth in SEQ ID NO: 175.

[1633] SEQ ID NO:1977 (encoded by SEQ ID NO: 11977) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and A1191L, P1195G.

[1634] SEQ ID NO:1978 (encoded by SEQ ID NO: 11978) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and P1195G.

[1635] SEQ ID NO:1163 (encoded by SEQ ID NO: 11163) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and L1200I.

[1636] SEQ ID NO:1164 (encoded by SEQ ID NO: 11164) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and N891S, T892V, S895V, E888S, E898S, T899V.

[1637] SEQ ID NO:1165 (encoded by SEQ ID NO: 11165) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and N891A, T892A, S895A, E888A, E898A, T899A.

[1638] SEQ ID NO:1166 (encoded by SEQ ID NO: 11166) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and T934V, G937P, P939G.

[1639] SEQ ID NO:1167 (encoded by SEQ ID NO: 11167) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and P593G, S595V, E596S, S599V, E888S, N891S, T892V, S895A.

[1640] SEQ ID NO:1168 (encoded by SEQ ID NO: 11168) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and P593G, S595V, E596S, S599V, E888S, T899A, N902S.

[1641] SEQ ID NO:1169 (encoded by SEQ ID NO: 11169) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and P593G, S595V, E596S, S599V, T934A, G937A.

[1642] SEQ ID NO:1170 (encoded by SEQ ID NO: 11170) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and N891S, T892V, S895V, K1252D, D1255K, W1256A.

[1643] FIGS. 51A and 51B are bar graphs showing results of integration assays using engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOs: 1163-1170 and 1977-1978. The engineered R2-TG drivers were used with the GFP reporter construct SEQ ID NO:175. The driver was added to the reporter after 3h or 6h. The drivers were engineered R2-TG protein constructs described above.

[1644] Most of the engineered R2-TG drivers described in this example exhibited improved activity compared to the WT R2-TG control (SEQ ID NO:1 +SEQ ID NO:175), except for SEQ ID NO: 1977 and SEQ ID NO:1170 which exhibited activity below WT level. The template for the amino acid substitutions in this experiment was SEQ ID NO:636 and the engineered R2-TG driver + WT R2-TG reporter (SEQ ID NO: 636 + SEQ ID NO:175) was one of the controls. SEQ ID NO:1163, SEQ ID NO:1164, SEQ ID NO:1165, SEQ ID NO:1166, SEQ ID NO:1167, SEQ ID NO:1168 and SEQ ID NO:1169 exhibited ~2-5 fold-improvement over the engineered R2-TG control (SEQ ID NO:636 +SEQ ID NO:175) when added 3h after the reporter, and similar levels or higher compared to their results when added 6h after the reporter. The activity that remained the same in the two time points are in contrast to the control which activity was reduced when added 6h after the reporter compared to addition 3h after the reporter. The improvement might be due to extended expression of the protein because of lower degradation rates.

Example 52

[1645] Driver constructs encoding engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOS: 764, 765, 770-774, 1321, 1322 and 1845 described, hereinbelow were generated and tested with the GFP reporter construct having a nucleic acid sequence set forth in SEQ ID NO: 175. SEQ ID NO:1323 sets forth the nucleic acid sequence of a modified reporter.

[1646] SEQ ID NO:764 (encoded by SEQ ID NO: 10764) sets forth the amino acid sequence of a R2 created by ancestral reconstruction process using R2 elements. The R2 element was reconstructed from monophyletic group 1 node 7 whose leaves are R2-MC SEQ ID NO: 411, R2-MC SEQ ID NO: 458, R2-GA SEQ ID NO: 408, R2-GA SEQ ID NO: 455, R2-NA SEQ ID NO: 409, R2-NA SEQ ID NO: 456 as well as HAG 5'UTR, HAG 3'UTR and 3 stop codons. The DNA encoding the polypeptide is codon-optimized using GenSmart™ from GenScript.

[1647] SEQ ID NO:765 (encoded by SEQ ID NO: 10765) sets forth the amino acid sequence of a R2 created by ancestral reconstruction process using R2 elements. The R2 element was reconstructed from monophyletic group 6 node 14 whose leaves are R2-OT SEQ ID NO:460, R2-OT SEQ ID NO:1336, R2-OT SEQ ID NO: 413 as well as HAG 5'UTR, HAG 3'UTR and 3 stop codons. The DNA encoding the polypeptide is codon-optimized using GenSmart™ from GenScript.

[1648] SEQ ID NO:770 (encoded by SEQ ID NO: 10770) sets forth the amino acid sequence of a R2 created by ancestral reconstruction process on R2 elements. The R2 element was reconstructed from monophyletic group 1 node 8 whose leaves are R2-GA SEQ ID NO: 408, R2-GA SEQ ID NO: 455, R2-NA SEQ ID NO: 409, R2-NA SEQ ID NO: 456 as well as HAG 5'UTR, HAG 3'UTR and 3 stop codons. The DNA encoding the polypeptide is codon-optimized using GenSmart™ from GenScript.

[1649] SEQ ID NO:771 (encoded by SEQ ID NO: 10771) sets forth the amino acid sequence of a R2 created by ancestral reconstruction process using R2 elements. The R2 element was reconstructed from monophyletic group 2 node 9 whose leaves are R2-XC SEQ ID NO: 474, R2-LS SEQ ID NO: 457, R2-LS SEQ ID NO: 410, R2-GP SEQ ID NO: 453, R2-GP SEQ ID NO: 407, R2-SM SEQ ID NO: 480 and R2-SM SEQ ID NO: 462 as well as HAG 5'UTR, HAG

3'UTR and 3 stop codons. The DNA encoding the polypeptide is codon-optimized using GenSmart™ from GenScript.

[1650] SEQ ID NO:774 (encoded by SEQ ID NO: 10774) sets forth the amino acid sequence of a R2 created by ancestral reconstruction process using R2 elements. The R2 element was reconstructed from monophyletic group 2 node 10 whose leaves are R2-LS SEQ ID NO: 457 and R2-LS SEQ ID NO: 410 as well as HAG 5'UTR, HAG 3'UTR and 3 stop codons. The DNA encoding the polypeptide is codon-optimized using GenSmart™ from GenScript.

[1651] SEQ ID NO:1845 (encoded by SEQ ID NO: 11845) sets forth the amino acid sequence of a R2 created by ancestral reconstruction process on R2 elements. The R2 element was reconstructed from monophyletic group 3 node 4 whose leaves are R2-OM SEQ ID NO:1367, R2-CC SEQ ID NO: 444, R2-CC SEQ ID NO: 402 as well as HAG 5'UTR, HAG 3'UTR and 3 stop codons. The DNA encoding the polypeptide is codon-optimized using GenSmart™ from GenScript.

[1652] SEQ ID NO:772 (encoded by SEQ ID NO: 10772) sets forth the amino acid sequence of a R2 created by ancestral reconstruction process using R2 elements. The R2 element was reconstructed from monophyletic group 6 node 10 whose leaves are R2-HN SEQ ID NO: 409, R2-HN SEQ ID NO: 456, R2-MC SEQ ID NO: 458, R2-MC SEQ ID NO: 411, R2-OT SEQ ID NO: 460, R2-OT SEQ ID NO: 1336, R2-OT SEQ ID NO: 413, R2-GA SEQ ID NO: 455 and R2-GA SEQ ID NO: 408 as well as HAG 5'UTR, HAG 3'UTR and 3 stop codons. The DNA encoding the polypeptide is codon-optimized using GenSmart™ from GenScript.

[1653] SEQ ID NO:773 (encoded by SEQ ID NO: 10773) sets forth the amino acid sequence of a R2 created by ancestral reconstruction process using R2 elements. The R2 element was reconstructed from monophyletic group 6 node 13 whose leaves are R2-OT SEQ ID NO:460, R2-OT SEQ ID NO:1336, R2-OT SEQ ID NO:413, R2-GA SEQ ID NO: 455 and R2-GA SEQ ID NO: 408 as well as HAG 5'UTR, HAG 3'UTR and 3 stop codons. The DNA encoding the polypeptide is codon-optimized using GenSmart™ from GenScript.

[1654] SEQ ID NO:1321 (encoded by SEQ ID NO: 11321) sets forth the amino acid sequence of a R2 created by ancestral reconstruction process using R2 elements. The R2 element was reconstructed from monophyletic group 7 node 5 whose leaves are R2-AC SEQ ID NO:440, R2-

AN SEQ ID NO:441, R2-SC SEQ ID NO:463, R2-ZA SEQ ID NO:1337 as well as HAG 5'UTR, HAG 3'UTR and 3 stop codons. The DNA encoding the polypeptide is codon-optimized using GenSmart™ from GenScript.

[1655] SEQ ID NO:1322 (encoded by SEQ ID NO: 11322) sets forth the amino acid sequence of a R2 created by ancestral reconstruction process using R2 elements. The R2 element was reconstructed from monophyletic group 8 node 6 whose leaves are R2-LS SEQ ID NO: 457, R2-LS SEQ ID NO: 410, R2-XC SEQ ID NO: 474, R2-GP SEQ ID NO: 407 and R2-GP SEQ ID NO: 453 as well as HAG 5'UTR, HAG 3'UTR and 3 stop codons. The DNA encoding the polypeptide is codon-optimized using GenSmart™ from GenScript.

[1656] The modified reporter construct is described, herein below:

[1657] SEQ ID NO:1323 (encoded by SEQ ID NO: 11323) sets forth GFP reporter comprising a transgene cassette (anti-sense, with CMV promoter and HSV TK polyA) with 5'UTR created by ancestral reconstruction process using R2 5'UTR sequences. The R2 5'UTR (SEQ ID NO:3008) was reconstructed from monophyletic group 5 node 4 whose leaves are R2-CD SEQ ID NO:1360, R2-CA SEQ ID NO:1361 and R2-CF SEQ ID NO:1362.

[1658] FIGS. 52A and 52B are bar graphs showing results of integration assays using engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOs: 764, 765, 770-774, 1321, 1322 and 1845 and one R2 reporter constructed by ancestor reconstruction process set forth in SEQ ID NO: 1323. In this experiment, R2 elements driver constructs generated by ancestor reconstructed process were used with the WT R2-TG GFP reporter construct SEQ ID NO:175 and a reporter created by ancestor reconstruction (SEQ ID NO:581) was tested with engineered R2-TG control driver (SEQ ID NO:636).

[1659] Transfection was carried out with driver and reporter mRNA delivered with lipofectamine to U2OS cells and full-length integration assessed by FACS at 24h. The drivers and reporter were products of ancestor reconstruction process done on active R2 elements.

[1660] The dataset included the following controls: Lipofectamine-only control; the following GFP reporter negative controls WT R2-TG control (SEQ ID NO:175), the reconstructed R2 reporter (SEQ ID NO:759) (WT R2-TG control driver with WT R2-TG reporter (SEQ ID NO:1

+SEQ ID NO:175) and engineered R2-TG control driver with WT R2-TG reporter (SEQ ID NO:636 +SEQ ID NO:175)

[1661] The reporter SEQ ID NO:1323 showed activity above that of the reference: WT R2-TG control driver with WT reporter (SEQ ID NO:1 + NO:175). While the engineered R2-TG with WT R2-TG reporter (SEQ ID NO:636 + SEQ ID NO:175) had ~3x fold improvement activity over WT reference reporter SEQ ID NO:1323 with the engineered control R2-TG driver has ~3.5x fold improvement of activity.

[1662] None of the reconstructed drivers were active.

Example 53

[1663] Driver constructs encoding engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOS: 1292, 1294-1298, 1417 and 1420 described, hereinbelow were generated and tested with the GFP reporter construct having a nucleic acid sequence set forth in SEQ ID NO: 581.

[1664] SEQ ID NO:1417 (encoded by SEQ ID NO: 11417) sets forth the amino acid sequence of a R2-PTT with HAG 5' UTR, HAG 3'UTR.

[1665] SEQ ID NO:1420 (encoded by SEQ ID NO: 11420) sets forth the amino acid sequence of a R2-LS with HAG 5'UTR, HAG 3'UTR and 3 stop codons.

[1666] SEQ ID NO: 408 sets forth the amino acid sequence of a R2-GA with HAG 5'UTR, HAG 3'UTR and 3 stop codons.

[1667] SEQ ID NO:1292 (encoded by SEQ ID NO: 11292) sets forth the amino acid sequence of a R2-chimera generated by R2-TG (SEQ ID NO: 1335, residues 1-474)+ R2-LS (SEQ ID NO:1420, residues 323-1239) with HAG 5'UTR, HAG 3'UTR.

[1668] SEQ ID NO:1294 (encoded by SEQ ID NO: 11294) sets forth the amino acid sequence of a R2-chimera generated by R2-LS (SEQ ID NO:1420, residues 1-322)+ R2-TG (SEQ ID NO:1293 residues 475-1390) with HAG 5'UTR, HAG 3'UTR.

[1669] SEQ ID NO:1295 (encoded by SEQ ID NO: 11295) sets forth the amino acid sequence of a R2-chimera generated by R2-LS (SEQ ID NO:1420, residues 1-84)+ R2-PTT (SEQ ID NO:1417, residues 148-287) + R2-LS (SEQ ID NO:1420 residues 208-292) + R2-TG (SEQ ID NO:1293, residues 459-1094) + R2-PTT (SEQ ID NO:1417 residues 1137-1434) with HAG 5'UTR, HAG 3'UTR.

[1670] SEQ ID NO:1296 (encoded by SEQ ID NO: 11296) sets forth the amino acid sequence of a R2-chimera generated R2-GA (SEQ ID NO: 408 residues 1-210) + R2-TG (SEQ ID NO:1293 residues 299-1290)+ R2-GA (SEQ ID NO: 408 residues 1162-1261) with HAG 5'UTR, HAG 3'UTR.

[1671] SEQ ID NO:1297 (encoded by SEQ ID NO: 11297) sets forth the amino acid sequence of a R2-chimera generated R2-GA (5 SEQ ID NO: 408 residues 1-210) + R2-PTT (SEQ ID NO:1417 residues 276-1334)+ R2-LS (SEQ ID NO:1420 residues 1141-1239) with HAG 5'UTR, HAG 3'UTR.

[1672] SEQ ID NO:1298 (encoded by SEQ ID NO: 11298) sets forth the amino acid sequence of a R2-chimera generated by R2-LS (SEQ ID NO:1420 (residues 1-93)+ R2-PTT (SEQ ID NO:1417 residues 161-436) + R2-TG (SEQ ID NO:1293 residues 459-1094) + R2-PTT (SEQ ID NO:1417 residues 1137-1434) with HAG 5'UTR, HAG 3'UTR.

[1673] FIGS. 53A and 53B are bar graphs showing results of integration assays using engineered R2-TG drivers comprising the combinations of retrotransposable elements R2-TG, R2-PTT, R2-LS, R2-GA and chimeras generated from these R2 RTEs. The engineered R2-TG drivers were used with the following GFP reporter constructs SEQ ID NO:581, R2-PTT SEQ ID NO:582 or R2-LS SEQ ID NO:758.

[1674] The dataset included the following controls: Lipofectamine-only control; GFP reporters SEQ ID NO:581, R2-PTT SEQ ID NO:582 and R2-LS SEQ ID NO:758 as negative controls and the standard positive control described, supra.

[1675] R2-PTT driver SEQ ID NO:1417 with R2-PTT reporter SEQ ID NO:582 showed increased activity compared to WT R2-TG control driver with modified reporter (SEQ ID NO:1 +SEQ ID NO:581).

[1676] One chimeric driver SEQ ID NO:1294 with R2 reporter SEQ ID NO:581 exhibited activity greater than the WT R2-TG control driver with modified reporter control (SEQ ID NO:1 +SEQ ID NO:581).

[1677] The rest of the chimera drivers with R2 reporter combinations showed reduced activity.

Example 54

[1678] Driver constructs encoding engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOS: 1299-1304 described, hereinbelow were generated and tested with the GFP reporter construct having a nucleic acid sequence set forth in SEQ ID NO:581, R2-PTT SEQ ID NO:582 or R2-LS SEQ ID NO:758

[1679] The dataset included the following controls: Lipofectamine™-only control; and GFP reporters SEQ ID NO:581, SEQ ID NO:582 and SEQ ID NO:758 as negative controls. The WT R2-TG control driver with modified reporter (SEQ ID NO:1 +SEQ ID NO:581 and the engineered R2-TG control driver with modified reporter (SEQ ID NO:636 +SEQ ID NO:581) as positive controls.

[1680] The driver constructs are described, herein below:

[1681] SEQ ID NO:1299 (encoded by SEQ ID NO: 11299) sets forth the amino acid sequence of a R2-chimera generated by R2-PTT (SEQ ID NO:1417, residues 1-287) + R2-LS (SEQ ID NO:1420, residues 208-292) + R2-TG (SEQ ID NO:1293 residues 459-1094) + R2-PTT (SEQ ID NO:1417 (residues 1137-1434) with HAG 5'UTR, HAG 3'UTR.

[1682] SEQ ID NO:1300 (encoded by SEQ ID NO: 11300) sets forth the amino acid sequence of a R2-chimera generated by R2-TG (SEQ ID NO:1293 residues 1-180) + R2-PTT (SEQ ID NO:1417 residues 148-287) + R2-LS (SEQ ID NO:1420 residues 208-292) + R2-TG (SEQ ID NO:1293 residues 459-1094) + R2-PTT (SEQ ID NO:1417, residues 1137-1434) with HAG 5'UTR, HAG 3'UTR.

[1683] SEQ ID NO:1301 (encoded by SEQ ID NO: 11301) sets forth the amino acid sequence of a R2-chimera generated by R2-LS (SEQ ID NO:1420 residues 1-84)+ R2-PTT (SEQ ID NO:1417 residues 148-436) + R2-TG (SEQ ID NO:1293 residues 459-1094) + R2-PTT (SEQ ID NO:1417 residues 1137-1434) with HAG 5'UTR, HAG 3'UTR.

[1684] SEQ ID NO:1302 (encoded by SEQ ID NO: 11302) sets forth the amino acid sequence of a R2-chimera generated by R2-TG (SEQ ID NO:1293 residues 1-180)+ R2-PTT (SEQ ID NO:1417 residues 148-436) + R2-TG (SEQ ID NO:1293 residues 459-1094) + R2-PTT (SEQ ID NO:1417 residues 1137-1434) with HAG 5'UTR, HAG 3'UTR.

[1685] SEQ ID NO:1303 (encoded by SEQ ID NO: 11303) sets forth the amino acid sequence of a R2-chimera generated by R2-LS (SEQ ID NO:1420 residues 1-84) + R2-PTT (SEQ ID NO:1417 residues 148-287) + R2-LS (SEQ ID NO:1420 residues 208-292) + R2-TG (SEQ ID NO:1293, residues 459-1390) with HAG 5'UTR, HAG 3'UTR.

[1686] SEQ ID NO:1304 (encoded by SEQ ID NO: 11304) sets forth the amino acid sequence of a R2-chimera generated by R2-TG (SEQ ID NO:1293, residues 1-206)+ R2-PTT (SEQ ID NO:1417 (residues 173-275) + R2-GA (SEQ ID NO: 408 residues 211-1161) + R2-TG (SEQ ID NO:1293 residues 1291-1390) with HAG 5'UTR, HAG 3'UTR.

[1687] FIGS. 54A and 54 B are bar graphs showing results of integration assays using engineered R2-TG drivers comprising the different retrotransposable chimera elements composed from R2-TG, R2-PTT, R2-LS, R2-GA (SEQ ID NO:1293, SEQ ID NO:1417, SEQ ID NO:1420 and SEQ ID NOs: 1335, 471, 410, 408. In this experiment, different retrotransposable chimera driver constructs were used with the GFP reporter constructs SEQ ID NO:581, R2-PTT SEQ ID NO:582 or R2-LS SEQ ID NO:758. The drivers were the R2-TG, R2-PTT, R2-LS and R2-GA R2-chimera protein constructs described above.

[1688] All chimera drivers with R2 reporter combinations showed reduced to abolished activity compared to the WT R2-TG control driver with modified reporter control (SEQ ID NO: +SEQ ID NO:581).

Example 55

[1689] Driver constructs encoding engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOS: 1305-1308 described, hereinbelow were generated and tested with the GFP reporter construct having a nucleic acid sequence set forth in SEQ ID NO: 581.

[1690] The driver constructs are described, herein below:

SEQ ID NO:1305 (encoded by SEQ ID NO: 11305) sets forth the amino acid sequence of a R2

chimeric polypeptide generated by R2-PTT (SEQ ID NO:1417 residues 1-436) + R2-TG (SEQ ID NO:1293 residues 459-1094) + R2-PTT (SEQ ID NO:1417 (residues 1137-1434) with HAG 5'UTR, HAG 3'UTR.

[1691] SEQ ID NO:1306 (encoded by SEQ ID NO: 11306) sets forth the amino acid sequence of a R2-chimera generated by R2-PTT (SEQ ID NO:1417 residues 1-287) + R2-LS (SEQ ID NO:1420 residues 208-292) + R2-TG (SEQ ID NO:1293 residues 459-1390) with HAG 5'UTR, HAG 3'UTR.

[1692] SEQ ID NO:1307 (encoded by SEQ ID NO: 11307) sets forth the amino acid sequence of a R2-chimera generated by R2-TG (SEQ ID NO:1293 residues 1-206)+ R2-PTT (SEQ ID NO:1417 residues 173-275) + R2-GA (SEQ ID NO: 408, residues 211-1161) + R2-PTT (SEQ ID NO:1417, residues 1335-1434) with HAG 5'UTR, HAG 3'UTR.

[1693] SEQ ID NO:1308 (encoded by SEQ ID NO: 11308) sets forth the amino acid sequence of a R2-chimera generated by R2-PTT (SEQ ID NO:1417 residues 1-275) + R2-GA (SEQ ID NO: 408 residues 211-1161) + R2-TG (SEQ ID NO:1293, residues 1291-1390) with HAG 5'UTR, HAG 3'UTR.

[1694] FIGS. 55A and 55B are bar graphs showing results of integration assays using engineered R2-TG drivers comprising the different retrotransposable chimera elements (SEQ ID NOS 1305-1308) composed from R2-TG, R2-PTT, R2-LS, R2-GA (SEQ ID NO:1293, SEQ ID NO:1417, SEQ ID NO:1420 SEQ ID NOS: 1335, 471, 410, 408). In this experiment, different retrotransposable chimera driver constructs were used with the GFP reporter constructs SEQ ID NO:581, R2-PTT SEQ ID NO:582 or R2-LS SEQ ID NO:758. The drivers were the R2-TG, R2-PTT, R2-LS and R2-GA R2-chimera protein constructs described above.

[1695] The dataset included the following controls: Lipofectamine™-only control; GFP reporters SEQ ID NO:581, R2-PTT SEQ ID NO:582 and R2-LS SEQ ID NO:758 as negative controls, and WT R2-TG control driver with modified reporter (SEQ ID NO:1 +SEQ ID NO:581) and R2-TG control driver with modified reporter (SEQ ID NO:636 +SEQ ID NO:581) as positive controls.

[1696] All chimera driver with R2 reporter combinations showed reduced to abolished activity compared to the WT R2-TG control driver with modified reporter control (SEQ ID NO:1 +SEQ ID NO:581).

Example 56

[1697] Driver constructs encoding engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOS: 1324-1327 and reporter construct 1328 described, hereinbelow were generated and tested.

[1698] The dataset included the standard controls.

[1699] SEQ ID NO: 1324 (encoded by SEQ ID NO: 11324) sets forth the amino acid sequence of a R2-chimera from R2-PTT (SEQ ID NO:1417 residues 1-147)+ R2-TG (SEQ ID NO:1293 residues 181-458) + R2-PTT (SEQ ID NO:1417, residues 437-1136)+ R2-TG (SEQ ID NO:1293 residues 1094-1390) with HAG 5'UTR, HAG 3'UTR and 3 stop codons.

[1700] SEQ ID NO: 1325 (encoded by SEQ ID NO: 5) sets forth the amino acid sequence of a R2-TG variant #1 found in R2-TG genome with HAG 5'UTR, HAG 3'UTR and 3 stop codons. The DNA encoding the polypeptide was codon optimized for fast translation codons

[1701] SEQ ID NO: 1326 (encoded by SEQ ID NO: 11326) sets forth the amino acid sequence of a R2-TG variant #5 found in R2-TG genome with HAG 5'UTR, HAG 3'UTR and 3 stop codons. The DNA encoding the polypeptide is codon-optimized using was codon optimized for fast translation codons.

[1702] SEQ ID NO: 1327 (encoded by SEQ ID NO: 11327) sets forth the amino acid sequence of a R2 created by ancestral reconstruction process using R2-TG variant #1 SEQ ID NO:1325, variant #2 SEQ ID NO:766, variant #3 SEQ ID NO:1315,variant #4 SEQ ID NO:767, variant #5 SEQ ID NO:1326 and WT R2-TG SEQ ID NO:1 as well as HAG5'UTR, HAG 3'UTR and 3 stop codons. The DNA encoding the polypeptide is codon-optimized using GenSmart™ from GenScript.

[1703] The reporter constructs are described, herein below:

[1704] SEQ ID NO:1328 sets forth the nucleic acid sequence of a GFP reporter comprising a transgene cassette (anti-sense, with CMV promoter and HSV TK polyA) with R2-TG 3'UTR variant #1.

[1705] SEQ ID NO:1329 sets forth the nucleic acid sequence of a GFP reporter comprising a transgene cassette (anti-sense, with CMV promoter and HSV TK polyA) with R2-TG 3'UTR variant #2.

[1706] SEQ ID NO:1330 sets forth the nucleic acid sequence of a GFP reporter comprising a transgene cassette (anti-sense, with CMV promoter and HSV TK polyA) with R2-TG 3'UTR variant #4.

[1707] SEQ ID NO:1331 sets forth the nucleic acid sequence of a GFP reporter comprising a transgene cassette (anti-sense, with CMV promoter and HSV TK polyA) with R2-TG 5'UTR variant #2.

[1708] SEQ ID NO:1332 sets forth the nucleic acid sequence of a GFP reporter comprising a transgene cassette (anti-sense, with CMV promoter and HSV TK polyA) with R2-TG 5'UTR variant #3.

[1709] SEQ ID NO:1333 sets forth the nucleic acid sequence of a GFP reporter comprising a transgene cassette (anti-sense, with CMV promoter and HSV TK polyA) with R2-TG 3'UTR variant #4.

[1710] SEQ ID NO:582 sets forth the nucleic acid sequence of a GFP reporter comprising a transgene cassette (anti-sense, with CMV promoter and HSV TK polyA) with R2-PTT natural UTRs.

[1711] SEQ ID NO:758 sets forth the nucleic acid sequence of a GFP reporter comprising a transgene cassette (anti-sense, with CMV promoter and HSV TK polyA) with R2-LS natural UTRs.

[1712] SEQ ID NO:1334) sets forth the nucleic acid sequence of a GFP reporter comprising a transgene cassette (anti-sense, with CMV promoter and HSV TK polyA) with R2-GFo natural UTRs.

[1713] FIGS. 56A and 56B are bar graphs showing results of integration assays using RTE drivers created by ancestor reconstruction process on R2-TG variants, and a chimera built from active R2 elements in addition to transgene cassette reporters for R2-PTT, R2-LS and R2-GFo and reporters constructed UTRs from R2-TG variants. In this experiment, R2-TG variant driver constructs and the R2-TG generated by ancestor reconstructed process were used with the R2-TG GFP reporter construct SEQ ID NO:581. The chimera was tested with R2-TG GFP reporter construct SEQ ID NO:581 or R2-PTT GFP reporter construct SEQ ID NO:582. All the tested reporters were checked with R2-TG engineered control driver SEQ ID NO:636.

[1714] When tested with the engineered control R2-TG driver SEQ ID NO:636 several reporters showed activity above that of the WT R2-TG control (SEQ ID NO:1 +SEQ ID NO:175). These reporters are: SEQ ID NO:1330, SEQ ID NO:1331, SEQ ID NO: 1332, SEQ ID NO:1333, SEQ ID NO:582 and SEQ ID NO:1334. The chimeric driver (SEQ ID NO:1324 in combination with R2-PTT SEQ ID NO:582 reporter also showed activity above that of the reference mentioned above. The engineered R2-TG driver with modified R2 reporter control (SEQ ID NO:636 + SEQ ID NO:581) exhibited ~2.7x fold improvement in activity over WT reference reporter. Three reporters had even higher improvement rates SEQ ID NO:1333, SEQ ID NO:582 and SEQ ID NO:1334 with ~3.8x fold, ~3.9x fold and ~3.4x fold improvement.

[1715] The R2-TG “variants” described in this and other examples refer to R2-TG retroelements that were found in the TG genome and differ by several mutations with respect to the WT R2-TG polypeptide set forth in SEQ IN NO: 1.

Example 57

[1716] Driver constructs encoding engineered R2-TG drivers having amino acid sequences described hereinbelow were generated and tested with the GFP reporter construct having a nucleic acid sequence set forth in SEQ ID NO: 1309-1312.

[1717] SEQ ID NO:1309 (encoded by SEQ ID NO: 11309) sets forth the amino acid sequence of a R2-chimera generated by R2-LS (SEQ ID NO:1420, residues 1-93)+ R2-TG (SEQ ID NO:1293 residues 192-458) + R2-PTT (SEQ ID NO:1417, residues 437-1136) + R2-TG (SEQ ID NO:1293 residues 1095-1390) with HAG 5'UTR, HAG 3'UTR and 3 stop codons.

[1718] SEQ ID NO:1310 (encoded by SEQ ID NO: 11310) sets forth the amino acid sequence of a R2-chimera with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide generated by R2-LS (SEQ ID NO:1420, residues 72-90)+ R2-PTT (SEQ ID NO:1417, residues 158-419) + R2-TG (SEQ ID NO:636 residues 442-1290) + R2-MC (SEQ ID NO: 458, residues 1140-1238) with HAG 5'UTR, HAG 3'UTR and 3 stop codons.

[1719] SEQ ID NO:1311 (encoded by SEQ ID NO: 11311) sets forth the amino acid sequence of a R2-chimera with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide generated by R2-LS (SEQ ID NO:1420, residues 73-207)+ R2-TG (SEQ ID NO:636, residues 311-433) + R2-GP (SEQ ID NO: 453, residues 290-1238) with HAG 5'UTR, HAG 3'UTR and 3 stop codons.

[1720] SEQ ID NO:1312 (encoded by SEQ ID NO: 11312) sets forth the amino acid sequence of a R2-chimera with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide generated by R2-LS (SEQ ID NO:1420, residues 73-207)+ R2-MC (SEQ ID NO: 458, residues 1140-1238) + R2-GP (SEQ ID NO: 453, residues 290-1238) with HAG 5'UTR, HAG 3'UTR and 3 stop codons.

[1721] The reporter construct is described, herein below:

[1722] SEQ ID NO:1873 sets forth GFP reporter comprising a transgene cassette (anti-sense, with CMV promoter and HSV TK polyA) with 3'UTRs R2-TG variant 5 from R2-TG polypeptide.

[1723] FIGS. 57A and 57B are bar graphs showing results of integration assays using chimeric RTE drivers composed from R2-TG, R2-LS, R2-GP and R2-MC (SEQ ID NOs: 1335, 636, 410, 453) and transgene cassette reporter with 3'UTR created by ancestral reconstruction process using 3'UTRs from active R2 elements (SEQ ID NO:1873).

[1724] In this experiment, chimeric R2 driver constructs were used with the GFP reporter constructs SEQ ID NO:581, SEQ ID NO:175, R2-PTT SEQ ID NO:582, R2-LS SEQ ID NO:758, R2-GP SEQ ID NO:759 or R2-MC SEQ ID NO: 3002. The drivers were R2-chimeras constructed from R2-TG, R2-LS, R2-GP and R2-MC and a tested reporter with 3'UTR from R2-TG variant.

[1725] The dataset included the following controls: Lipofectamine™-only and the GFP reporters as negative controls, and standard positive controls described supra.

[1726] Several chimera drivers with reporter combinations improved activity above that of WT R2-TG control (SEQ ID NO:1 +SEQ ID NO:175): SEQ ID NO:1309 with either SEQ ID NO:581 or SEQ ID NO:582 reporter; SEQ ID NO:1311 with either SEQ ID NO:581 SEQ ID NO:759; SEQ ID NO:1312 with SEQ ID NO:759. Chimera SEQ ID NO:1311 with SEQ ID NO:758 has the same activity as the WT R2-TG control.

[1727] The activity of the reporter (SEQ ID NO:1873) with control R2-TG driver SEQ ID NO:636 was the same as 5'UTR modified reporter SEQ ID NO:595 + SEQ ID NO:636.

Example 58

[1728] Driver constructs encoding engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOS: 1314, 1313 described, hereinbelow were generated and tested with the GFP reporter construct having a nucleic acid sequence set forth in SEQ ID NO: 581, 582, 100759, 175.

[1729] The driver constructs are described, herein below:

[1730] SEQ ID NO:1314 (encoded by SEQ ID NO: 11314) sets forth the amino acid sequence of a R2-chimera with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide generated by R2-GP (SEQ ID NO: 453 residues 72-207)+ R2-PTT (SEQ ID NO:1417 residues 288-411) + R2-XC (SEQ ID NO: 474 residues 291-1146) + R2-TG (SEQ ID NO:636 residues 1291-1390) with HAG 5'UTR, HAG 3'UTR and 3 stop codons.

[1731] SEQ ID NO:1313 (encoded by SEQ ID NO: 11313) sets forth the amino acid sequence of a R2-chimera with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide generated by R2-GP (SEQ ID NO: 453 residues 72-207)+ R2-PTT (SEQ ID NO:1417 residues 288-411) + R2-XC (SEQ ID NO: 474 residues 291-1245) with HAG 5'UTR, HAG 3'UTR and 3 stop codons.

[1732] FIGS. 58A and 58B are bar graphs showing results of integration assays using chimeric RTE drivers composed from R2-TG, R2-PTT, R2-GP and R2-XC SEQ ID NOs: 636, 1417, 453, 474). In this experiment, chimera R2 elements driver constructs were tested with the GFP

reporters having AA sequences set forth in (SEQ ID NO:581), R2-PTT (SEQ ID NO:582), R2-GP (SEQ ID NO:759) or R2-XC (SEQ ID NO: 3001)

[1733] The drivers were R2-chimera built from R2-TG, R2-LS, R2-GP and R2-XC. The dataset included the following controls: Lipofectamine™-only and the following GFP reporter negative controls: (SEQ ID NO:581), wildtype R2-TG (SEQ ID NO:175), R2-PTT (SEQ ID NO:582) R2-GP (SEQ ID NO:759) and R2-XC (SEQ ID NO: 3001) as negative controls and the standard positive controls described supra.

[1734] Both of the chimera drivers with reporter combinations improved activity above that of WT R2-TG control (SEQ ID NO:1 +SEQ ID NO:581) The engineered R2-TG control (SEQ ID NO:636) exhibited ~2x fold improvement activity over WT. Most chimera driver with R2 reporter combinations showed better activity than the engineered SEQ ID NO:636 with activity ranging between ~3x and ~5x fold higher. These combinations are SEQ ID NO:1314 with one of reporters SEQ ID NO:581, SEQ ID NO:582 SEQ ID NO:759; or SEQ ID NO:1313 with reporters SEQ ID NO:582, SEQ ID NO:759 or SEQ ID NO:3001.

Example 59

[1735] Driver constructs encoding engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOS: 1315-1317 and 1294 described, hereinbelow were generated and tested with the GFP reporter construct having a nucleic acid sequence set forth in SEQ ID NO: 581,582.

[1736] The driver constructs are described, herein below:

[1737] SEQ ID NO:1315 (encoded by SEQ ID NO: 11315) sets forth the amino acid sequence of a R2-TG variant #3 found in R2-TG genome with HAG 5'UTR, HAG 3'UTR and 3 stop codons. The DNA encoding the polypeptide was codon optimized for fast translation codons.

[1738] SEQ ID NO:1316 (encoded by SEQ ID NO: 11316) sets forth the amino acid sequence of a R2-chimera with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide generated by R2-XC (SEQ ID NO: 474 residues 72-208)+ R2-PTT (SEQ ID NO:1417, residues 288-411) + R2-LS (SEQ ID NO:1420, residues 290-1140) + R2-PTT (SEQ ID NO:1417, residues 1336-1434) with HAG 5'UTR, HAG 3'UTR and 3 stop codons.

[1739] SEQ ID NO:1317 (encoded by SEQ ID NO: 11317) sets forth the amino acid sequence of a R2-chimera with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide generated by R2-LS (SEQ ID NO:1420 residues 73-207)+ R2-PTT (SEQ ID NO:1417 residues 288-411) + R2-XC (SEQ ID NO:474 residues 291-1146) + R2-MC (SEQ ID NO: 458 residues 1160-1430) with HAG 5'UTR, HAG 3'UTR and 3 stop codons.

[1740] SEQ ID NO:1294 (encoded by SEQ ID NO: 11294) sets forth the amino acid sequence of a R2-chimera generated by R2-LS (SEQ ID NO:1420 residues 1-322)+ R2-TG (SEQ ID NO:1293 residues 475-1390) with HAG 5'UTR, HAG 3'UTR.

[1741] FIGS. 59A and 59B are bar graph showing results of integration assays using chimeric RTE drivers composed from R2-TG, R2-PTT, R2-LS, R2-MC and R2-XC species and engineered R2-TG driver SEQ ID NO:636. In this experiment, the engineered R2-TG drivers were tested with the GFP reporters SEQ ID NO:581, SEQ ID NO:582, SEQ ID NO:758, SEQ ID NO: 3002 or SEQ ID NO: 3001.

[1742] The dataset included the following controls: Lipofectamine™-only and the following GFP reporter negative controls SEQ ID NO:581, SEQ ID NO:582, SEQ ID NO:758, SEQ ID NO: 3002 or SEQ ID NO: 3001 and the standard positive controls described supra.

[1743] Two chimera driver elements and the R2-TG variant 3 showed activity above that of WT R2-TG control (SEQ ID NO:1 +SEQ ID NO:175). The engineered R2-TG control (SEQ ID NO:636 + SEQ ID NO:581) exhibited a ~4x fold improvement activity over the WT control. When combined with R2-PTT reporter (SEQ ID NO:582) SEQ ID NO:636 had ~7x fold increase in activity. SEQ ID NO:636 with R2-MC reporter having SEQ ID NO: 3002 shows better activity than the R2-TG control driver with modified reporter but not from SEQ ID NO:636 +SEQ ID NO:581).

[1744] Chimera SEQ ID NO:1294 with R2-PTT reporter SEQ ID NO:582 showed ~5x fold increase in activity with respect to WT R2-TG control driver with WT R2 control (SEQ ID NO:1 +SEQ ID NO:175).

Example 60

[1745] Driver constructs encoding engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOS: 3027, 1318, 1319 and 1320 described, hereinbelow were generated and tested with the GFP reporter construct having a nucleic acid sequence set forth in SEQ ID NO: 175.

[1746] The driver constructs are described, herein below:

[1747] SEQ ID NO: 3027 (encoded by SEQ ID NO:13028) sets forth the amino acid sequence of a R2-chimera from R2-TG (SEQ ID NO:1293 residues 1-458)+ R2-PTT (SEQ ID NO:1417 residues 437-1136) + R2-TG (SEQ ID NO:1293 residues 1095-1390) with HAG 5'UTR, HAG 3'UTR and 3 stop codons.

[1748] SEQ ID NO:1318 (encoded by SEQ ID NO: 11318) sets forth the amino acid sequence of a R2-chimera with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide generated by R2-GP (SEQ ID NO: 453 residues 72-207)+ R2-MC (SEQ ID NO: 458 residues 221-302) + R2-XC (SEQ ID NO: 474 residues 291-1146) + R2-MC (SEQ ID NO: 458 residues 1160-1430) with HAG 5'UTR, HAG 3'UTR and 3 stop codons.

[1749] SEQ ID NO:1319 (encoded by SEQ ID NO: 11319) sets forth the amino acid sequence of a R2-chimera with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide generated by R2-GP (SEQ ID NO: 453 residues 72-207)+ R2-MC (SEQ ID NO: 458 residues 221-302) + R2-XC (SEQ ID NO: 474 residues 291-1146) + R2-TG (SEQ ID NO:636, residues 1291-1390) with HAG 5'UTR, HAG 3'UTR and 3 stop codons.

[1750] SEQ ID NO:1320 (encoded by SEQ ID NO: 11320) sets forth the amino acid sequence of a R2-chimera with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide generated by R2-GP (SEQ ID NO: 453 residues 72-207)+ R2-MC (SEQ ID NO: 458 residues 221-302) + R2-XC (SEQ ID NO: 474 residues 291-1245) with HAG 5'UTR, HAG 3'UTR and 3 stop codons.

[1751] FIGS. 60A and 60B are bar graphs showing results of integration assays using chimeric RTE drivers composed from R2-TG, R2-PTT, R2-MC, R2-GP and R2-XC set forth in SEQ ID

NOs: 3027, 1318, 1319 and 1320. In this experiment, R2 chimera element driver constructs were used with the GFP reporter constructs SEQ ID NO:581, SEQ ID NO:582, SEQ ID NO: 175.

[1752] The dataset included the following controls: Lipofectamine™-only and GFP reporters SEQ ID NO:581, SEQ ID NO:175, SEQ ID NO:582, and SEQ ID NO: 3001 as negative controls and the standard positive controls, described supra.

[1753] Six chimera driver - reporter combinations showed activity above that of the reference: WT R2-TG control (SEQ ID NO:1 + SEQ ID NO:175). These combinations are SEQ ID NO: 3027 + SEQ ID NO:582, SEQ ID NO:1318 + SEQ ID NO:582, SEQ ID NO:1318 + SEQ ID NO:3001, SEQ ID NO:1319 + SEQ ID NO:582, SEQ ID NO:1319 + SEQ ID NO:3001 and SEQ ID NO:11320 + SEQ ID NO:582.

[1754] The engineered R2-TG control (SEQ ID NO:636 + SEQ ID NO:581) had ~3x fold improvement activity over WT R2-TG control. Two chimera combinations (SEQ ID NO:1319 + SEQ ID NO:582) and (SEQ ID NO:1320 + SEQ ID NO:582) showed higher increase with 3.3x and 5.2x fold respectively over WT reference.

Example 61

[1755] Driver constructs encoding engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOS:1299-1304 described hereinbelow were generated and tested with the GFP reporter constructs having a nucleic acid sequence set forth in SEQ ID NO:581 and SEQ ID NO:582.

[1756] The driver constructs are described, herein below:

[1757] SEQ ID NO:1299 (encoded by SEQ ID NO: 11299) sets forth the amino acid sequence of a R2-chimera comprising sections from R2-PTT (SEQ ID NO:1417 residues 1-287) + R2-LS (SEQ ID NO:1420 residues 208-292) + R2-TG (SEQ ID NO:1293 residues 459-1094) + R2-PTT (SEQ ID NO:1417 residues 1137-1434) with HAG 5'UTR, HAG 3'UTR and 3 stop codons.

[1758] SEQ ID NO:1300 (encoded by SEQ ID NO: 11300) sets forth the amino acid sequence of a R2-chimera comprising sections from R2-TG (SEQ ID NO:1293 residues 1-180)+ R2-PTT (SEQ ID NO:1417 residues 148-287) + R2-LS (SEQ ID NO:1420 residues 208-292) + R2-TG

(SEQ ID NO:1293 residues 459-1094) + R2-PTT (SEQ ID NO:1417 (residues 1137-1434) with HAG 5'UTR, HAG 3'UTR and 3 stop codons.

[1759] SEQ ID NO:1301 (encoded by SEQ ID NO: 11301) sets forth the amino acid sequence of a R2-chimera comprising sections from R2-LS (SEQ ID NO:1420 residues 1-84)+ R2-PTT (SEQ ID NO:1417 residues 148-436) + R2-TG (SEQ ID NO:1293 residues 459-1094) + R2-PTT (SEQ ID NO:1417 residues 1137-1434) with HAG 5'UTR, HAG 3'UTR and 3 stop codons.

[1760] SEQ ID NO:1302 (encoded by SEQ ID NO: 11302) sets forth the amino acid sequence of a R2-chimera comprising sections from R2-TG (SEQ ID NO:1293 residues 1-180)+ R2-PTT (SEQ ID NO:1417 residues 148-436) + R2-TG (SEQ ID NO:1293 residues 459-1094) + R2-PTT (SEQ ID NO:1417 residues 1137-1434) with HAG 5'UTR, HAG 3'UTR and 3 stop codons.

[1761] SEQ ID NO:1303 (encoded by SEQ ID NO: 11303) sets forth the amino acid sequence of a R2-chimera comprising sections from R2-LS (SEQ ID NO:1420 residues 1-84)+ R2-PTT (SEQ ID NO:1417 residues 148-287) + R2-LS (SEQ ID NO:1420 residues 208-292) + R2-TG (SEQ ID NO:1293 residues 459-1390) with HAG 5'UTR, HAG 3'UTR and 3 stop codons.

[1762] SEQ ID NO:1304 (encoded by SEQ ID NO: 11304) sets forth the amino acid sequence of a R2-chimera comprising sections from R2-TG (SEQ ID NO:1293 residues 1-206)+ R2-PTT (SEQ ID NO:1417 residues 173-275) + R2-GA (SEQ ID NO: 408 residues 211-1161)+ R2-TG (SEQ ID NO:1293 residues 1291-1390) with HAG 5'UTR, HAG 3'UTR and 3 stop codons.

[1763] FIGS. 61A and 61B are bar graphs showing results of integration assays using chimeric RTE drivers composed from R2-TG, R2-PTT, R2-LS and R2-GA. In this experiment, the R2 chimera driver constructs were used with the GFP reporter constructs SEQ ID NO:581 SEQ ID NO:582.

[1764] The dataset included the following controls: Lipofectamine™-only and the GFP reporters SEQ ID NO:581, SEQ ID NO:582 SEQ ID NO:758 as negative controls, and the standard positive controls described supra.

[1765] None of the chimera driver with R2 reporter combinations showed activity above that of the reference: WT R2-TG control (SEQ ID NO:1 +SEQ ID NO:581).

Example 62

[1766] Driver constructs encoding engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOS: 1293, 1375, 1390, 1397, 1398, 1422, 1596, 1597, 1636, 1712, 1725-1734 described, hereinbelow were generated and tested with the GFP reporter construct having a nucleic acid sequence set forth in SEQ ID NO: 175.

[1767] SEQ ID NO:1390 (encoded by SEQ ID NO: 11390) sets forth the amino acid sequence of an R2-TG polypeptide having the K1307A amino acid substitution.

[1768] SEQ ID NO:1725 (encoded by SEQ ID NO: 11725) sets forth the amino acid sequence of an R2-TG polypeptide having the amino acid substitutions F148K, P280K, E399M, D555K, D923K, T1011S, I1219N, A1282G, V661K.

[1769] SEQ ID NO:1726 (encoded by SEQ ID NO: 11726) sets forth the amino acid sequence of an R2-TG polypeptide having the amino acid substitutions F148K, P280K, E399M, D555K, D923K, T1011S, I1219N, A1282G, C518Y.

[1770] SEQ ID NO:1727 (encoded by SEQ ID NO: 11727) sets forth the amino acid sequence of an R2-TG polypeptide having the amino acid substitutions F148K, P280K, E399M, D555K, D923K, T1011S, I1219N, A1282G, K700D.

[1771] SEQ ID NO:1728 (encoded by SEQ ID NO: 11728) sets forth the amino acid sequence of an R2-TG polypeptide having the amino acid substitutions F148K, P280K, E399M, D555K, D923K, T1011S, I1219N, A1282G, F667G.

[1772] SEQ ID NO:1729 (encoded by SEQ ID NO: 11729) sets forth the amino acid sequence of an R2-TG polypeptide having the amino acid substitutions F148K, P280K, E399M, D555K, D923K, T1011S, I1219N, A1282G, C145G.

[1773] SEQ ID NO:1730 (encoded by SEQ ID NO: 11730) sets forth the amino acid sequence of an R2-TG polypeptide having the includes amino acid substitutions F148K, P280K, E399M, D555K, D923K, T1011S, I1219N, A1282G, N889G.

[1774] SEQ ID NO:1731 (encoded by SEQ ID NO: 11731) sets forth the amino acid sequence of an R2-TG polypeptide having the amino acid substitutions F148K, P280K, E399M, D555K, D923K, T1011S, I1219N, A1282G, E399M.

[1775] SEQ ID NO:1732 (encoded by SEQ ID NO: 11732) sets forth the amino acid sequence of an R2-TG polypeptide having the amino acid substitutions F148K, P280K, E399M, D555K, D923K, T1011S, I1219N, A1282G, D923K.

[1776] SEQ ID NO:1733 (encoded by SEQ ID NO: 11733) sets forth the amino acid sequence of an R2-TG polypeptide having the amino acid substitutions F148K, P280K, E399M, D555K, D923K, T1011S, I1219N, A1282G, E399M, C518Y.

[1777] SEQ ID NO:1734 (encoded by SEQ ID NO: 11734) sets forth the amino acid sequence of an R2-TG polypeptide having the amino acid substitutions F148K, P280K, E399M, D555K, D923K, T1011S, I1219N, A1282G, N889G, C518Y.

[1778] SEQ ID NO:1596 (encoded by SEQ ID NO: 11596) sets forth the amino acid sequence of an R2-TG polypeptide having the amino acid substitutions F148K, P280K, E399M, D555K, D923K, T1011S, I1219N, A1282G, E214R.

[1779] SEQ ID NO:1597 (encoded by SEQ ID NO: 11597) sets forth the amino acid sequence of an R2-TG polypeptide chimera, with the native RT domain replaced by a Marathon RT and further includes amino acid substitutions F148K, P280K, E399M, D555K, D923K, T1011S, I1219N, A1282G.

[1780] SEQ ID NO:1636 (encoded by SEQ ID NO: 11636) sets forth the amino acid sequence of an R2-TG polypeptide chimera, with the native RT domain replaced by a R2-TOC RT and further includes amino acid substitutions F148K, P280K, E399M, D555K, D923K, T1011S, I1219N, A1282G.

[1781] SEQ ID NO:1712 (encoded by SEQ ID NO: 11712) sets forth the amino acid sequence of a R2-TG polypeptide with C-terminal fusion of an RNA binding peptide and further includes amino acid substitutions F148K, P280K, D555K, T1011S, I1219N, A1282G.

[1782] SEQ ID NO:1397 (encoded by SEQ ID NO: 11397) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue E184) with N-terminal

fusion of a SV40 NLS, HMGN1 polypeptide and with C-terminal fusion of an RNA binding peptide and further includes amino acid substitutions P280K, D555K.

[1783] SEQ ID NO:1398 (encoded by SEQ ID NO: 11398) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue E184) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and with C-terminal fusion of an RNA binding peptide and further includes amino acid substitutions P280K, D555K.

[1784] SEQ ID NO:1293 (encoded by SEQ ID NO: 11293) sets forth the amino acid sequence of an R2-TG polypeptide. R2-TG polypeptide includes amino acid substitutions F148K, P280K, E399M, K700D, D555K, D923K, T1011S, I1219N, A1282G.

[1785] SEQ ID NO:1422 (encoded by SEQ ID NO: 11422) sets forth the amino acid sequence of an R2-TG polypeptide having the amino acid substitutions F148K, P280K, E399M, D555K, D923K, T1011S, I1219N, A1282G.

[1786] SEQ ID NO:1375 (encoded by SEQ ID NO: 11375) sets forth the amino acid sequence of an R2-TG polypeptide having the D555K amino acid substitution.

[1787] FIGS. 62A- 62B are bar graphs showing results of integration assays using engineered R2-TG drivers amino acid substitution having an amino acid sequence set forth in SEQ ID NOS: SEQ ID NOS: 1293, 1375, 1390, 1397, 1398, 1422, 1596, 1597, 1636, 1712, 1725-1734. Transfection was carried out with driver and reporter mRNA delivered with Lipofectamine™ to U2OS cells and full-length integration assessed by FACS at 24h. The drivers were engineered R2-TG protein constructs described above.

[1788] The dataset included the following controls: Lipofectamine™-only and GFP reporter (SEQ ID NO:175) as negative controls; WT RT-TG control (SEQ ID NO: 1 + SEQ ID NO: 175) and Vingi trans element (SEQ ID NO: 1716 + SEQ ID NO: 175) as positive controls. Error bars indicate standard error of two replicates.

[1789] Most of the engineered drivers exhibited improved or similar activity to SEQ ID NO: WT R2-TG control (SEQ ID NO:1 +SEQ ID NO:175). The improved constructs had 20-230% improved activity compared to the WT R2-TG control. SEQ ID NO:1390 with a K1307A amino acid substitution seemingly had a negative effect on the endonuclease domain activity. SEQ ID

NO:1597 with the Marathon RT substitution also had lower activity than the WT control, possibly due to a negative effect on the protein structure or folding.

Example 63: R2-TG drivers having truncated N-termini exhibit improved activity

[1790] Driver constructs encoding engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOS: 1473, 1477, 1519 and 1735-1745 described, hereinbelow were generated and tested with the GFP reporter construct having a nucleic acid sequence set forth in SEQ ID NO: 175.

[1791] ID NO:1735 (encoded by SEQ ID NO: 11735) sets forth the amino acid sequence of an R2-TG polypeptide having the amino acid substitutions G257R, D555K, C145G, E214R, L213K, D537Y, A1282G, P280K, F148K, K700D, F667G, Y1242K, T989L, E542K, N889G, D923K.

[1792] SEQ ID NO:1736 (encoded by SEQ ID NO: 11736) sets forth the amino acid sequence of an R2-TG polypeptide having the amino acid substitutions D1133P, R669L, D537Y, T1011A, F148K, F973E, K1085R, K547RR, N363H, A1282G, G257R.

[1793] SEQ ID NO:1737 (encoded by SEQ ID NO: 11737) sets forth the amino acid sequence of an R2-TG polypeptide having the amino acid substitutions D555K, P280K, D1133P, R669L, D537Y, T1011A, F148K, F973E, K1085R, K547R, N363H, A1282G.

[1794] SEQ ID NO:1738 (encoded by SEQ ID NO: 11738) sets forth the amino acid sequence of an R2-TG polypeptide with N-terminal fusion of a SV40 NLS, paRecT polypeptide and further includes amino acid substitutions D1133P, R669L, D537Y, T1011A, F148K, F973E, K1085R, K547R, N363H, A1282G, G257R, A442P, E399M.

[1795] SEQ ID NO:1739 (encoded by SEQ ID NO: 11739) sets forth the amino acid sequence of an R2-TG polypeptide having the amino acid substitutions C145G, F148K, E214R, P280K, E399M, D555K, K700D, S895K, D923K, T989L, T1011S, I1219N, Y1242K, A1282G.

[1796] SEQ ID NO:1740 (encoded by SEQ ID NO: 11740) sets forth the amino acid sequence of an R2-TG polypeptide chimera, with the native RT domain replaced by a Marathon RT and further includes amino acid substitutions P280K, D555K.

[1797] SEQ ID NO:1741 (encoded by SEQ ID NO: 11741) sets forth the amino acid sequence of an R2-TG polypeptide chimera, with the native RT domain replaced by a Marathon RT and further includes amino acid substitutions C145G, D555K, A1012G.

[1798] SEQ ID NO:1742 (encoded by SEQ ID NO: 11742) sets forth the amino acid sequence of an R2-TG polypeptide chimera, with the native RT domain replaced by a RTX RT and further includes amino acid substitutions P280K, D555K.

[1799] SEQ ID NO:1743 (encoded by SEQ ID NO: 11743) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, Sto7D polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, T989L.

[1800] SEQ ID NO:1744 (encoded by SEQ ID NO: 11744) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue E184) with N-terminal fusion of a SV40 NLS, Sto7D polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, T989L.

[1801] SEQ ID NO:1745 (encoded by SEQ ID NO: 11745) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue E184) with N-terminal fusion of a SV40 NLS, Sto7D polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K.

[1802] SEQ ID NO:1473 (encoded by SEQ ID NO: 11473) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, Sto7D polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, K700D.

[1803] SEQ ID NO:1477 (encoded by SEQ ID NO: 11477) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, Sto7D polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, T989L.

[1804] SEQ ID NO:1519 (encoded by SEQ ID NO: 11519) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal

fusion of a SV40 NLS, CTIP polypeptide, Sto7D polypeptide. R2-TG polypeptide includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, E214R.

[1805] SEQ ID NO:630 (encoded by SEQ ID NO: 10630) has been described, supra in Example 17.

[1806] SEQ ID NO:631 (encoded by SEQ ID NO: 10631) has been described, supra in Example 17.

[1807] SEQ ID NO:636 (encoded by SEQ ID NO: 10636) has been described, supra.

[1808] FIGS. 63A and 63B are bar graphs showing results of integration assays using engineered R2-TG drivers amino acid substitution having an amino acid sequence set forth in SEQ ID NOs: 1473, 1477, 1519 and 1735-1745 in Example 17.

[1809] Transfection was carried out with driver and reporter mRNA delivered with Lipofectamine™ to U2OS cells and full-length integration assessed by FACS at 24h. The drivers were engineered R2-TG protein constructs described above.

[1810] The dataset included the following controls: Lipofectamine™-only and GFP reporter (SEQ ID NO:175) as negative controls; WT RT-TG control (SEQ ID NO: 1 + SEQ ID NO: 175) and Vingi trans element (SEQ ID NO: 1716 + SEQ ID NO: 175) as positive controls. Error bars indicate standard error of two replicates.

[1811] Most of the engineered drivers with amino acid substitutions or combination of modifications had improved or similar activity to WT R2-TG control (SEQ ID NO:1 +SEQ ID NO:175). The improved constructs had 20-300% improved activity compared to the WT R2-TG control.

[1812] SEQ ID NO:1740, SEQ ID NO:1741, SEQ ID NO:1742 with the Marathon or RTX RT substitution had lower activity than the WT, probably due to negative effect on the protein structure or folding.

Example 64

[1813] Driver constructs encoding engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOS: 1390, 1743, and 1750-1764 described, hereinbelow were generated and tested with the GFP reporter construct having a nucleic acid sequence set forth in SEQ ID NO: 175.

[1814] SEQ ID NO:1390 (encoded by SEQ ID NO: 11390) sets forth the amino acid sequence of an R2-TG polypeptide having the K1307A amino acid substitution.

[1815] SEQ ID NO:1743 (encoded by SEQ ID NO: 11743) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, Sto7D polypeptide, CTIP polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, T989L.

[1816] SEQ ID NO:1750 (encoded by SEQ ID NO: 11750) sets forth the amino acid sequence of an R2-TG polypeptide having the amino acid substitutions F148K, P280K, D555K, T1011S, I1219N, T989L.

[1817] SEQ ID NO:1751 (encoded by SEQ ID NO: 11751) sets forth the amino acid sequence of an R2-TG polypeptide having the amino acid substitutions F148K, P280K, D555K, T1011S, I1219N, V660R.

[1818] SEQ ID NO:1752 (encoded by SEQ ID NO: 11752) sets forth the amino acid sequence of an R2-TG polypeptide having the amino acid substitutions F148K, P280K, D555K, T1011S, I1219N, E638L.

[1819] SEQ ID NO:1753 (encoded by SEQ ID NO: 11753) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, HMGN1 polypeptide, Sto7D polypeptide further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K.

[1820] SEQ ID NO:1754 (encoded by SEQ ID NO: 11754) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue E184) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, Sto7D polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, E214R.

[1821] SEQ ID NO:1755 (encoded by SEQ ID NO: 11755) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, HMGN1 polypeptide, CTIP polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, T989L.

[1822] SEQ ID NO:1756 (encoded by SEQ ID NO: 11756) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, HMGN1 polypeptide, CTIP polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K.

[1823] SEQ ID NO:1757 (encoded by SEQ ID NO: 11757) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, HMGN1 polypeptide, CTIP polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, E214R.

[1824] SEQ ID NO:1758 (encoded by SEQ ID NO: 11758) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, K700D.

[1825] SEQ ID NO:1759 (encoded by SEQ ID NO: 11759) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, E214R.

[1826] SEQ ID NO:1760 (encoded by SEQ ID NO: 11760) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, RAD51 polypeptide, CTIP polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G. The DNA encoding the RT domain of the R2-TG polypeptide is codon-optimized for the most frequent codons.

[1827] SEQ ID NO:1761 (encoded by SEQ ID NO: 11761) sets forth the amino acid sequence of an R2-TG polypeptide having the amino acid substitutions F148K, P280K, E399M, D555K, D923K, T989L, T1011S, I1219N.

[1828] SEQ ID NO:1762 (encoded by SEQ ID NO: 11762) sets forth the amino acid sequence of an R2-TG polypeptide having the amino acid substitutions F148K, P280K, E399M, D555K, D923K, K579R, T1011S, I1219N.

[1829] SEQ ID NO:1763 (encoded by SEQ ID NO: 11763) sets forth the amino acid sequence of an R2-TG polypeptide having the amino acid substitutions F148K, P280K, E399M, D555K, D923K, G257R, T1011S, I1219N.

[1830] SEQ ID NO:1764 (encoded by SEQ ID NO: 11764) sets forth the amino acid sequence of an R2-TG polypeptide having the amino acid substitutions F148K, P280K, E399M, D555K, D923K, F876Y, T1011S, I1219N.

[1831] SEQ ID NO:629 (encoded by SEQ ID NO: 10629) has been described, supra in Example 17.

[1832] SEQ ID NO:636 (encoded by SEQ ID NO: 10636) has been described, supra

[1833] FIGS. 64A and 64B are bar graphs showing results of integration assays using engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOs: 1390, 1743, and 1750-1764. Drivers having amino acid sequences set forth in any one of SEQ ID NOS: 1750, 1751, 1752, 1753, 1754, 1755, 1756, 1757, 1758, 1759, 1760, 1761, 1762, 1763 exhibited approximately 1.5-3 fold increase in activity over the WT R2-TG control (SEQ ID NO: 1 + SEQ ID NO: 175). SEQ ID NO: 1758 exhibited activity similar to the engineered R2-TG control (SEQ ID NO:636+ SEQ ID NO:175).

[1834] Transfection was carried out with driver and reporter mRNA delivered with Lipofectamine™ to U2OS cells and full-length integration assessed by FACS at 24h. The drivers were engineered R2-TG protein constructs described above.

[1835] The dataset included the following controls: Lipofectamine™-only and GFP reporter (SEQ ID NO:175) as negative controls; WT RT-TG control (SEQ ID NO: 1 + SEQ ID NO: 175) and Vingi trans element (SEQ ID NO: 1716 + SEQ ID NO: 175) as positive controls. Error bars indicate standard error of two replicates.

[1836] Most of the engineered drivers with amino acid substitutions or combination of modifications had similar activity or improved activity by 50-400% compared to WT R2-TG control (SEQ ID NO:1 +SEQ ID NO:175).

[1837] SEQ ID NO:1390 with a K1307A amino acid substitution seemingly had a negative effect on the endonuclease domain activity.

Example 65

[1838] Driver constructs encoding engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOS: 1781-1786 and 1788-1793 described, hereinbelow were generated and tested with the GFP reporter construct having a nucleic acid sequence set forth in SEQ ID NO: 175.

[1839] SEQ ID NO:1390 (encoded by SEQ ID NO: 11390) sets forth the amino acid sequence of an R2-TG polypeptide having the K1307A amino acid substitution.

[1840] SEQ ID NO:1781 (encoded by SEQ ID NO: 11781) sets forth the amino acid sequence of an R2-TG polypeptide having amino acid substitutions F148K, P280K, D555K, T1011S, I1219N, A1282G, K700D, C518Y, D923K.

[1841] SEQ ID NO:1782 (encoded by SEQ ID NO: 11782) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, Sto7D polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, E214R.

[1842] SEQ ID NO:1783 (encoded by SEQ ID NO: 11783) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue E184) with N-terminal fusion of a SV40 NLS, Sto7D polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, E214R.

[1843] SEQ ID NO:1784 (encoded by SEQ ID NO: 11784) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, Sto7D polypeptide, CTIP polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, G257R.

[1844] SEQ ID NO:1785 (encoded by SEQ ID NO: 11785) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, Sto7D polypeptide, CTIP polypeptide. R2-TG polypeptide includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, E399M.

[1845] SEQ ID NO:1786 (encoded by SEQ ID NO: 11786) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, HMGN1 polypeptide, CTIP polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, K700D.

[1846] SEQ ID NO:1788 (encoded by SEQ ID NO: 11788) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, HMGN1 polypeptide, CTIP polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K, C518Y.

[1847] SEQ ID NO:1789 (encoded by SEQ ID NO: 11789) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue E184) with N-terminal fusion of a SV40 NLS, HMGN1 polypeptide, Sto7D polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, G257R.

[1848] SEQ ID NO:1790 (encoded by SEQ ID NO: 11790) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue E184) with N-terminal fusion of a SV40 NLS, Sto7D polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, G257R.

[1849] SEQ ID NO:1791 (encoded by SEQ ID NO: 11791) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, Sto7D polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, G257R.

[1850] SEQ ID NO:1792 (encoded by SEQ ID NO: 11792) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, Sto7D polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, E399M.

[1851] SEQ ID NO:1793 (encoded by SEQ ID NO: 11793) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue E184) with N-terminal fusion of a SV40 NLS, Sto7D polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, E399M.

[1852] SEQ ID NO:625 (encoded by SEQ ID NO: 10625) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue E184) with N-terminal fusion of a SV40 NLS, HMGN1 polypeptide, Sto7D polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, K700D has been described, supra in Example 17.

[1853] SEQ ID NO:627 (encoded by SEQ ID NO: 10627) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue E184) with N-terminal fusion of a SV40 NLS, Sto7D polypeptide, HMGN1 polypeptide. R2-TG polypeptide includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, K700D has been described, supra in Example 17.

[1854] SEQ ID NO:628 (encoded by SEQ ID NO: 10628) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, Sto7D polypeptide. R2-TG polypeptide includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K has been described, supra in Example 17.

[1855] SEQ ID NO:636 (encoded by SEQ ID NO: 10636) has been described, supra.

[1856] SEQ ID NO:715 (encoded by SEQ ID NO: 10715) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide. R2-TG polypeptide includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K has been described, supra in Example 14.

[1857] FIGS. 65A and 65B are bar graph showing results of integration assays using engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOs: 1781-1786 and 1788-1793.

[1858] Transfection was carried out with driver and reporter mRNA delivered with Lipofectamine™ to U2OS cells and full-length integration assessed by FACS at 24h. The drivers were engineered R2-TG protein constructs described above.

[1859] The dataset included the following controls: Lipofectamine™-only and GFP reporter (SEQ ID NO:175) as negative controls; WT RT-TG control (SEQ ID NO: 1 + SEQ ID NO: 175) and Vingi trans element (SEQ ID NO: 1716 + SEQ ID NO: 175) as positive controls. Error bars indicate standard error of two replicates.

[1860] Most of the engineered drivers with amino acid substitutions or combination of modifications had improved or similar activity to WT R2-TG control (SEQ ID NO:1 +SEQ ID NO:175). The improved constructs had 20-350% improved activity compared to the WT R2-TG.

Example 66

[1861] Driver constructs encoding engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOS: 1218 and 1220-1232 described hereinbelow were generated and tested with the GFP reporter construct having a nucleic acid sequence set forth in SEQ ID NO:175.

[1862] SEQ ID NO:1218 (encoded by SEQ ID NO: 11218) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of SV40 NLS, HMGN1 polypeptide, CTIP polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G.

[1863] SEQ ID NO:1220 (encoded by SEQ ID NO: 11220) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue H147) with N-terminal fusion of HMGN1 polypeptide and further includes the amino acid substitutions F148K, P280K, E399M, D555K, D923K, T1011S, I1219N, A1282G.

[1864] SEQ ID NO:1221 (encoded by SEQ ID NO: 11221) forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of SV40 NLS and CTIP polypeptide and C-terminal fusion of RNA binding motif and further includes amino acid substitutions D555K, P280K, T1011S.

[1865] SEQ ID NO:1222 (encoded by SEQ ID NO: 11222) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal

fusion of SV40 NLS, HMGN1 and CTIP polypeptide and C-terminal fusion of RNA binding motif and further includes amino acid substitutions D555K, P280K.

[1866] SEQ ID NO:1223 (encoded by SEQ ID NO: 11223) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue E184) with N-terminal fusion of HMGN1 polypeptide and C-terminal fusion of RNA binding motif and further includes amino acid substitutions D555K, P280K, T1011S.

[1867] SEQ ID NO:1224 (encoded by SEQ ID NO: 11224) sets forth the amino acid sequence of a R2-TG with N-terminal fusion of PARP1 NLS and further includes amino acid substitutions F148K, P280K, E399M, D555K, D923K, T1011S, I1219N, A1282G.

[1868] SEQ ID NO:1225 (encoded by SEQ ID NO: 11225) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of SV40 NLS and CTIP polypeptide. The RT domain in the R2-TG was replaced with the RT domain from RT-TOC. R2-TG polypeptide and further includes amino acid substitutions D555K, P280K, T1011S, I1219N.

[1869] SEQ ID NO:1226 (encoded by SEQ ID NO: 11226) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of SV40 NLS and HMGN1 polypeptide. The RT domain in the R2-TG was replaced with the RT domain from RT-TOC. R2-TG polypeptide and further includes amino acid substitutions D555K, P280K, T1011S.

[1870] SEQ ID NO:1227 (encoded by SEQ ID NO: 11227) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue E184) with N-terminal fusion of SV40 NLS and HMGN1 polypeptide. The RT domain in the R2-TG was replaced with the RT domain from RT-TOC. R2-TG polypeptide and further includes amino acid substitutions D555K, P280K, T1011S.

[1871] SEQ ID NO:1228 (encoded by SEQ ID NO: 11228) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue E184) with N-terminal fusion of SV40 NLS CTIP and HMGN1 polypeptide. The RT domain in the R2-TG was replaced with the RT domain from RT-TOC. R2-TG polypeptide includes amino acid substitutions D555K, P280K.

[1872] SEQ ID NO:1229 (encoded by SEQ ID NO: 11229) sets forth the amino acid sequence of a R2-TG whose RT domain were replaced with the RT-Marathon. R2-TG polypeptide includes amino acid substitutions D555K, P280K, F148K, T1011S.

[1873] SEQ ID NO:1230 (encoded by SEQ ID NO: 11230) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of SV40 NLS and CTIP. The RT domain in the R2-TG was replaced with R2-Marathon. R2-TG polypeptide includes amino acid substitutions D555K, P280K, T1011S.

[1874] SEQ ID NO:1231 (encoded by SEQ ID NO: 11231) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue E184) with N-terminal fusion of HMGN1 polypeptide. The RT domain in the R2-TG was replaced with R2-Marathon. R2-TG polypeptide includes amino acid substitutions D555K, P280K.

[1875] SEQ ID NO:1232 (encoded by SEQ ID NO: 11232) sets forth amino acid sequence of R2-TG with C-terminal RNA binding motif and further includes amino acid substitutions D555K, P280K, F148K, T1011S.

[1876] SEQ ID NO:1422 (encoded by SEQ ID NO: 11422) has been described, supra, in Example 63.

[1877] SEQ ID NO:618 (encoded by SEQ ID NO: 10618) has been described supra, in Example 17.

[1878] SEQ ID NO:635 (encoded by SEQ ID NO: 10635) has been described supra in Example 17.

[1879] SEQ ID NO:636 (encoded by SEQ ID NO: 10636) has been described, supra.

[1880] FIGS. 66A and 66B are bar graphs showing results for drivers with multiple domain additions in combination with improving amino acid substitutions, having amino acid sequences set forth in SEQ ID NOS: 1218 and 1220-1232.

[1881] Transfection was carried out with driver and reporter mRNA delivered with Lipofectamine™ to U2OS cells and full-length integration assessed by FACS at 24h. The drivers were engineered R2-TG protein constructs described above.

[1882] The dataset included the following controls: Lipofectamine™-only and GFP reporter (SEQ ID NO:175) as negative controls; WT RT-TG control (SEQ ID NO: 1 + SEQ ID NO: 175), Vingi trans element (SEQ ID NO: 1716 + SEQ ID NO: 175) and another engineered R2-TG control (SEQ ID NO:1422+ SEQ ID NO: 175) as positive controls. Error bars indicate standard error of two replicates.

[1883] Most constructs showed improved the activity with activity above that of WT R2-TG (SEQ ID NO:1 +SEQ ID NO:175). Nine constructs showed improved the activity above that of another engineered R2-TG control (SEQ ID NO:1422+SEQ ID NO:175). These are SEQ ID NO:635, SEQ ID NO:1218, SEQ ID NO:636, SEQ ID NO:1221, SEQ ID NO:1222, SEQ ID NO:1224, SEQ ID NO:1225, SEQ ID NO:1226 (and SEQ ID NO: 1228 (encoded by SEQ ID NO: 11228)).The improvement was in the range of 12% (SEQ ID NO:635) to 120% (SEQ ID NO:636).

Example 67

[1884] Driver constructs encoding engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOS: 619, 1686 and 1688-1701 described, hereinbelow were generated and tested with the GFP reporter construct having a nucleic acid sequence set forth in SEQ ID NO: 175.

[1885] The driver constructs are described, herein below:

[1886] SEQ ID NO:1686 (encoded by SEQ ID NO: 11686) sets forth amino acid sequence of R2-TG with amino acid substitutions P280K, F148K, D555K, K700D, F667G, Y1242K, T989L, C145G, E542K, N889G, D923K, C518Y, E630D, L316R.

[1887] SEQ ID NO:1688 (encoded by SEQ ID NO: 11688) sets forth amino acid sequence of R2-TG with amino acid substitutions D555K, P280K, D1133P, R669L, D537Y, T1011A, F148K, F973E, K1085R, K547R.

[1888] SEQ ID NO:1689 (encoded by SEQ ID NO: 11689) sets forth the amino acid sequence of a R2-TG with N-terminal fusion of RecT from *pseudomonas aeruginosa* (paRecT) polypeptide and further includes amino acid substitutions D1133P, R669L, D537Y, T1011A, F148K, F973E, K1085R, K547R, N363H, A1282G.

[1889] SEQ ID NO:1690 (encoded by SEQ ID NO: 11690) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of SV40 NLS and CTIP polypeptide and further includes amino acid substitutions P280K, E399M, D555K, D923K, T1011S, I1219N, A1282G.

[1890] SEQ ID NO:1691 (encoded by SEQ ID NO: 11691) sets forth amino acid sequence of R2-TG with amino acid substitutions C145G, F148K, P280K, E399M, D555K, D923K, T1011S, I1219N, A1282G.

[1891] SEQ ID NO:1692 (encoded by SEQ ID NO: 11692) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue E184) with N-terminal fusion of SV40 NLS and CTIP and HMGN1 polypeptides in addition to C-terminal fusion of RNA binding motif and further includes amino acid substitutions D555K, P280K, T1011S.

[1892] SEQ ID NO:1693 (encoded by SEQ ID NO: 11693) sets forth amino acid sequence of R2-TG with amino acid substitutions L51G, F148K, P280K, E399M, D555K, D923K, T1011S, I1219N, A1282G.

[1893] SEQ ID NO:1694 (encoded by SEQ ID NO: 11694) sets forth amino acid sequence of R2-TG with amino acid substitutions F148K, L213R, P280K, E399M, D555K, D923K, T1011S, I1219N, A1282G.

[1894] SEQ ID NO:1695 (encoded by SEQ ID NO: 11695) sets forth the amino acid sequence with N-terminal fusion of PARP1 NLS and further includes amino acid substitutions F148K, P280K, E399M, C518Y, D555K, K700D, D923K, T1011S, I1219N, A1282G.

[1895] SEQ ID NO:1696 (encoded by SEQ ID NO: 11696) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of SV40 NLS and CTIP polypeptide and further includes amino acid substitutions E214R, P280K, E399M, D555K, D923K, T989L, T1011S, I1219N, A1282G.

[1896] SEQ ID NO:1697 (encoded by SEQ ID NO: 11697) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue E184) with N-terminal fusion of HMGN1 polypeptide and CTIP polypeptide and further includes amino acid substitutions F148K, P280K, E399M, D555K, D923K, T1011S, I1219N, A1282G.

[1897] SEQ ID NO:1698 (encoded by SEQ ID NO: 11698) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of SV40 NLS, CTIP polypeptide and STO7D polypeptide.

[1898] SEQ ID NO:1699 (encoded by SEQ ID NO: 11699) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue E184) with N-terminal fusion of HMGN1 polypeptide and STO7D polypeptide.

[1899] SEQ ID NO:1700 (encoded by SEQ ID NO: 11700) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of SV40 NLS and BRCA2 peptide and further includes amino acid substitutions P280K, E399M, D555K, D923K, T1011S, I1219N, A1282G.

[1900] SEQ ID NO:1701 (encoded by SEQ ID NO: 11701) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of SV40 NLS, HMGN1 polypeptide and CTIP polypeptide and C-terminal fusion of RNA binding motif and further and further includes amino acid substitutions D555K, P280K, T1011S.

[1901] SEQ ID NO:619 (encoded by SEQ ID NO: 10619) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue E184) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, E399M, D555K, D923K, T1011S, I1219N, A1282G.

[1902] SEQ ID NO:1293 (encoded by SEQ ID NO: 11293) sets forth amino acid sequence of R2-TG with amino acid substitutions F148K, P280K, E399M, D555K, K700D, D923K, T989L, T1011S, I1219N, A1282G.

[1903] SEQ ID NO:1422 (encoded by SEQ ID NO: 11422) sets forth amino acid sequence of R2-TG with amino acid substitutions F148K, P280K, E399M, D555K, D923K, T1011S, I1219N, A1282G.

[1904] FIGS. 67A and 67B are bar graphs showing results for drivers with multiple domain additions in combination with top of improving amino acid substitutions found previously (SEQ ID NO: 619, 1686 and 1688-1701).

[1905] Transfection was carried out with driver and reporter mRNA delivered with Lipofectamine™ to U2OS cells and full-length integration assessed by FACS at 24h. The drivers were the R2 constructs described above.

[1906] The dataset included the following controls: Lipofectamine™-only and Dead R2-TG driver (SEQ ID NO:1390 + SEQ ID NO:175) as negative controls; and WT RT-TG control (SEQ ID NO:1 + SEQ ID NO:175), an engineered RT-TG control driver with WT reporter (SEQ ID NO:1422 + SEQ ID NO:175) as positive controls. A positive control for high integration SEQ ID NO:1716 (encoded by SEQ ID NO: 11716) was further included.

[1907] Thirteen constructs showed improved the activity with activity above that of wild-type R2-TG control: SEQ ID NO:619, SEQ ID NO:1690, SEQ ID NO:1691, SEQ ID NO:1692, SEQ ID NO:1693, SEQ ID NO:1694, SEQ ID NO:1695, SEQ ID NO:1696, SEQ ID NO:1697, SEQ ID NO:1698, SEQ ID NO:1699, SEQ ID NO:1701, SEQ ID NO:1422.

[1908] Seven constructs showed improved activity above that of control engineered R2-TG (SEQ ID NO:1422+SEQ ID NO:175): SEQ ID NO:1690, SEQ ID NO:1693, SEQ ID NO:1696, SEQ ID NO:619, SEQ ID NO:1697 SEQ ID NO:1698, SEQ ID NO 1701. The improvement was in the range of 13% (SEQ ID NO:1698) to 110% (SEQ ID NO:1690 and SEQ ID NO:619)

Example 68

[1909] Driver constructs encoding engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOS: 1423-1439 described, hereinbelow were generated and tested with the GFP reporter construct having a nucleic acid sequence set forth in SEQ ID NO: 175.

[1910] The driver constructs are described, herein below:

[1911] SEQ ID NO:1423 (encoded by SEQ ID NO: 11423) sets forth the amino acid sequence of a R2-TG with amino acid substitutions P280K, F148K, D555K, K700D, F667G, Y1242K, T989L, C145G, E542K, N889G, D923K, C518Y, E630D, L316R, E400M, Y571H, D1075E, M671A, E214R, L213K, T1016L, E350R, F667Y, P180K.

[1912] SEQ ID NO:1424 (encoded by SEQ ID NO: 11424) sets forth the amino acid sequence of a R2-TG with amino acid substitutions G257R, D555K, C145G, E214R, L213K, D537Y, A1282G, P280K, F148K, K700D, F667G, Y1242K, T989L, E542K, N889G, D923K, C518Y, E630D, L316R, E400M, Y571H

[1913] SEQ ID NO:1425 (encoded by SEQ ID NO: 11425) sets forth the amino acid sequence of a R2-TG with N-terminal fusion of paRecT polypeptide and further includes amino acid substitutions D1133P, R669L, D537Y, T1011A, F148K, F973E, K1085R, K547R, N363H, A1282G, G257R, A442P, E399M, V539G, R1177G, H423Q, D555K, N889G, Q1019R, Y1242K.

[1914] SEQ ID NO:1426 (encoded by SEQ ID NO: 11426) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide and HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, L564Q, Q567L, T1011S, I1219N, A1282G.

[1915] SEQ ID NO:1428 (encoded by SEQ ID NO: 11428) sets forth the amino acid sequence of a R2-TG including amino acid substitutions P280K, F148K, D555K, K700D, F667G, Y1242K, T989L, C145G, E542K, N889G, D923K, C518Y, E630D, L316R, E400M, Y571H, D1075E, M671A, E214R, L213K, T1016L, E350R, F667Y, P180K, P78V.

[1916] SEQ ID NO:1429 (encoded by SEQ ID NO: 11429) sets forth the amino acid sequence of a R2-TG including amino acid substitutions G257R, D555K, C145G, E214R, L213K, D537Y, A1282G, P280K, F148K, K700D, F667G, Y1242K, T989L, E542K, N889G D923K, C518Y, E630D, L316R, E400M, Y571H, D1075E

[1917] SEQ ID NO:1430 (encoded by SEQ ID NO: 11430) sets forth the amino acid sequence of a R2-TG including amino acid substitutions D1133P, R669L, D537Y, T1011A, F148K, F973E, K1085R, K547RR, N363H, A1282G, G257R, A442P, E399M, V539G, R1177G, H423Q, D555K, N889G.

[1918] SEQ ID NO:1431 (encoded by SEQ ID NO: 11431) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, UL42 polypeptide, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G. SEQ ID NO:1432

(encoded by SEQ ID NO: 11432) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, PF8 polypeptide, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G.

[1919] SEQ ID NO:1433 (encoded by SEQ ID NO: 11433) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and insertion of P21 PIP peptide between residues 563 and 564 and further includes amino acid substitutions P280K, D555K, Y571H, T1011S, I1219N, A1282G.

[1920] SEQ ID NO:1434 (encoded by SEQ ID NO: 11434) sets forth the amino acid sequence of a R2-TG including amino acid substitutions P280K, F148K, D555K, K700D, F667G, Y1242K, T989L, C145G, E542K, N889G, D923K, C518Y, E630D, L316R, E400M, Y571H, D1075E, M671A, E214R, L213K, T1016L, E350R, F667Y, P180K, P78V, R669I.

[1921] SEQ ID NO:1435 (encoded by SEQ ID NO: 11435) sets forth the amino acid sequence of a R2-TG including amino acid substitutions D1133P, R669L, D537Y, T1011A, F148K, F973E, K1085R, K547RR, N363H, A1282G, G257R, A442P, E399M, V539G, R1177G, H423Q, D555K, N889G, Q1019R, Y1242K

[1922] SEQ ID NO:1436 (encoded by SEQ ID NO: 11436) sets forth the amino acid sequence of a R2-TG including amino acid substitutions D555K, P280K, D1133P, R669L, D537Y, T1011A, F148K, F973E, K1085R, K547R, N363H, A1282G, G257R, A442P, E399M, V539G, R1177G.

[1923] SEQ ID NO:1437 (encoded by SEQ ID NO: 11437) sets forth the amino acid sequence of a R2-TG with N-terminal fusion of paRecT polypeptide and further includes amino acid substitutions D555K, P280K, D1133P, R669L, D537Y, T1011A, F148K, F973E, K1085R, K547R, N363H, A1282G, G257R, A442P, E399M, V539G.

[1924] SEQ ID NO:1438 (encoded by SEQ ID NO: 11438) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, paRecT polypeptide, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G.

[1925] SEQ ID NO:1439 (encoded by SEQ ID NO: 11439) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, EIF4A polypeptide, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G.

[1926] SEQ ID NO:1427 sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G and Y571H.

[1927] FIGS. 68A and 68B are bar graphs showing results for drivers with multiple domain additions in combination with improving amino acid substitutions found previously, having amino acid sequences set forth in SEQ ID NOS: 1423-1439.

[1928] Transfection was carried out with driver and reporter mRNA delivered with Lipofectamine™ to U2OS cells and full-length integration assessed by FACS at 24h. The drivers were the R2 constructs described above. The data set included the following controls: Lipofectamine™-only control, WT R2-TG GFP reporter SEQ ID NO:175) as negative controls: WT RT-TG control driver with WT reporter (SEQ ID NO: 10001 + SEQ ID NO: 175) and engineered RT-TG control driver with WT reporter control (SEQ ID NO: 100636 + SEQ ID NO: 175) positive controls and a positive control for high integration, SEQ ID NO: 101716.

[1929] Most of the engineered R2-TG drivers exhibited improved activity above that of WT R2-TG control: SEQ ID NO:1423, SEQ ID NO:1424, SEQ ID NO:1427, SEQ ID NO:1428, SEQ ID NO: 1439, SEQ ID NO:1431, SEQ ID NO:1432, SEQ ID NO:1437, SEQ ID NO: and SEQ ID NO:1426. SEQ ID NO:1427 includes N-terminal truncation of R2-TG up to V161, SV40 NLS, DNA repair recruitment peptide CTIP, chromatin binding domain HMGN1 and amino acid substitutions P280K, D555K, Y571H, T1011S, I1219N, A1282G, indicating an activity enhancing effect of amino acid substitution Y571H. SEQ ID NO:1431 includes the N-terminal truncation of R2-TG up to V161, viral processivity factor, SV40 NLS, UL42 polypeptide, DNA repair recruitment peptide CTIP, chromatin binding domain HMGN1 and amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G indicating the activity enhancing effect of UL42 polypeptide.

Example 69

[1930] Driver constructs encoding engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOS: 1394-1395, 1702-1711 and 1714-1715 described, hereinbelow were generated and tested with the GFP reporter construct having a nucleic acid sequence set forth in SEQ ID NO: 175.

[1931] The driver constructs are described, herein below:

[1932] SEQ ID NO:1702 (encoded by SEQ ID NO: 11702) sets forth the amino acid sequence of a R2-TG including amino acid substitutions P280K, F148K, D555K, K700D, F667G, Y1242K, T989L, C145G, E542K, N889G, D923K, C518Y, E630D, L316R, E400M.

[1933] SEQ ID NO:1703 (encoded by SEQ ID NO: 11703) sets forth the amino acid sequence of a R2-TG including amino acid substitutions P280K, F148K, D555K, K700D, F667G, Y1242K, T989L, C145G, E542K, N889G, D923K, C518Y, E630D, L316R, E400M, Y571H.

[1934] SEQ ID NO:1704 (encoded by SEQ ID NO: 11704) sets forth the amino acid sequence of a R2-TG including amino acid substitutions G257R, D555K, C145G, E214R, L213K, D537Y, A1282G, P280K, F148K, K700D, F667G, Y1242K, T989L, E542K.

[1935] SEQ ID NO:1705 (encoded by SEQ ID NO: 11705) sets forth the amino acid sequence of a R2-TG including amino acid substitutions D1133P, R669L, D537Y, T1011A, F148K, F973E, K1085R, K547RR, N363H.

[1936] SEQ ID NO:1706 (encoded by SEQ ID NO: 11706) sets forth the amino acid sequence of a R2-TG including amino acid substitutions D555K, P280K, D1133P, R669L, D537Y, T1011A, F148K, F973E, K1085R, K547R, N363H.

[1937] SEQ ID NO:1707 (encoded by SEQ ID NO: 11707) sets forth the amino acid sequence of a R2-TG with N-terminal fusion of paRecT polypeptide and further includes amino acid substitutions D1133P, R669L, D537Y, T1011A, F148K, F973E, K1085R, K547R, N363H, A1282G, G257R, A442P.

[1938] SEQ ID NO:1708 (encoded by SEQ ID NO: 11708) sets forth the amino acid sequence of a R2-TG with N-terminal fusion of paRecT polypeptide and further includes amino acid substitutions D555K, P280K, D1133P, R669L, D537Y, T1011A, F148K, F973E.

[1939] SEQ ID NO:1709 (encoded by SEQ ID NO: 11709) sets forth the amino acid sequence of a R2-TG with C-terminal fusion of RNA binding motif and further includes amino acid substitutions D555K, P280K, F148K, T1011S, I1219N, A1282G.

[1940] SEQ ID NO:1710 (encoded by SEQ ID NO: 11710) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40, NMGN1 polypeptide, CTIP polypeptide and C-terminal fusion of RNA binding motif and further includes amino acid substitutions D555K, P280K, T1011S, I1219N, A1282G.

[1941] SEQ ID NO:1711 (encoded by SEQ ID NO: 11711) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue E184) with N-terminal fusion of a SV40, NMGN1 polypeptide and C-terminal fusion of RNA binding motif and further includes amino acid substitutions D555K, P280K, T1011S, I1219N, A1282G, D700K.

[1942] SEQ ID NO:1394 (encoded by SEQ ID NO: 11394) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of SV40, CTIP polypeptide, NMGN1 polypeptide and C-terminal fusion of RNA binding motif and further includes amino acid substitutions D555K, P280K, T1011S, I1219N.

[1943] SEQ ID NO:1395 (encoded by SEQ ID NO: 11395) sets forth the amino acid sequence of a R2-TG including amino acid substitutions C145G, F148K, P280K, E399M, D555K, K700D, D923K, T989L, T1011S, I1219N, A1282G.

[1944] SEQ ID NO:1714 (encoded by SEQ ID NO: 11714) sets forth the amino acid sequence of a R2-TG including amino acid substitutions F148K, P280K, E399M, D555K, K700D, D923K, T1011S, I1219N, Y1242K, A1282G.

[1945] SEQ ID NO:1715 (encoded by SEQ ID NO: 11715) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue E184) with N-terminal

fusion of NMGN1 polypeptide. R2-TG polypeptide includes amino acid substitutions F148K, P280K, E399M, E542K, D555K, K579R, D923K, T1011S, I1219N, A1282G.

[1946] The reporter construct is described, herein below:

[1947] SEQ ID NO:614 sets forth a GFP transgene cassette (anti-sense, with CMV promoter and HSV TK polyadenylation sequence) with 3 repeats of the native R2-TG 5' RTE-UTR.

[1948] FIGS. 69A and 69B are bar graphs showing results for drivers with multiple domain additions in combination with improving amino acid substitutions having amino acid sequences set forth in any one of SEQ ID NOS: 1394-1385, 1702-1711 and 1714-1715. In this experiment, different retrotransposable element driver constructs were used with WT R2-TG GFP reporter construct SEQ ID NO:175 and a reporter (SEQ ID NO:614) was used with engineered T2-TG driver SEQ ID NO:636 and R2-ZA driver SEQ ID NO:1835

[1949] Transfection was carried out with driver and reporter mRNA delivered with Lipofectamine™ to U2OS cells and full-length integration assessed by FACS at 24h. The drivers and reporter were the R2 constructs described above. The dataset included the following controls: Lipofectamine™-only, WT R2-TG GFP reporter SEQ ID NO:175 with R2-TG dead driver SEQ ID NO:1390, engineered R2-TG driver SEQ ID NO:614 and R2-TG dead driver SEQ ID NO:1390 as negative controls. The WT RT-TG driver with WT reporter control (SEQ ID NO:1 +SEQ ID NO:175), engineered RT-TG control driver with WT reporter (SEQ ID NO:1378 +SEQ ID NO:175) as positive controls and a positive control for high integration SEQ ID NO:1716 (encoded by SEQ ID NO: 11716).

[1950] Most engineered R2-TG drivers exhibited improved activity above that of wild-type R2-TG SEQ ID NO: 1702, SEQ ID NO: 1703, SEQ ID NO:1704, SEQ ID NO:1705, SEQ ID NO:1706, SEQ ID NO:1709, SEQ ID NO:1710, SEQ ID NO:1711, SEQ ID NO:1394, SEQ ID NO:1395, SEQ ID NO:1714, SEQ ID NO:1715. In particular, SEQ ID NO:1710 which includes an N-terminal truncation of R2-TG up to V161, SV40 NLS, chromatin binding domain HMGN1, DNA repair recruitment peptide CTIP, C-terminal fusion of RNA binding motif and amino acid substitutions D555K, P280K, T1011S, I1219N, A1282G and SEQ ID NO:1395 which includes the following ten amino acid substitutions C145G, F148K, P280K, E399M, D555K, K700D, D923K, T989L, T1011S, I1219N, A1282G, also shows high activity with

respect to WT control but not higher than the positive control of engineered R2-TG SEQ ID NO:1378.

Example 70

[1951] Driver constructs encoding engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOS: 1765- and 3024 described, hereinbelow were generated and tested

[1952] The driver constructs are described, herein below:

[1953] SEQ ID NO: 3024 (encoded by SEQ ID NO: 13024) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue E184) with N-terminal fusion of bpNLS (bipartite SV40 NLS), HCV peptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N. The DNA encoding the polypeptide is codon-optimized for the most frequent codons.

[1954] SEQ ID NO:1765 (encoded by SEQ ID NO: 11765) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue E184) with N-terminal fusion of bpNLS (bipartite SV40 NLS), NSP3 peptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N. The DNA encoding the polypeptide is codon-optimized for the most frequent codons.

[1955] SEQ ID NO:1766 (encoded by SEQ ID NO: 11766) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue E184) with C-terminal fusion of NSP3 peptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N. The DNA encoding the polypeptide is codon-optimized for the most frequent codons.

[1956] SEQ ID NO:1767 (encoded by SEQ ID NO: 11767) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue E184) with N-terminal fusion of bpNLS (bipartite SV40 NLS), JFNS3 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N. The DNA encoding the polypeptide is codon-optimized for the most frequent codons.

[1957] SEQ ID NO:1768 (encoded by SEQ ID NO: 11768) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue E184) with N-terminal fusion of bpNLS (bipartite SV40 NLS), EshA polypeptide and further includes amino acid

substitutions P280K, D555K, T1011S, I1219N. The DNA encoding the polypeptide is codon-optimized for the most frequent codons.

[1958] SEQ ID NO:1769 (encoded by SEQ ID NO: 11769) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue E184) and further includes amino acid substitutions P280K, D555K, T1011S, I1219N. The DNA encoding the polypeptide is codon-optimized for the most frequent codons. The construct included an added 5'UTR HAG (SEQ ID NO: 3026).

[1959] SEQ ID NO:1770 (encoded by SEQ ID NO: 11770) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion SV40 NLS, BRACA2 polypeptide, CTIP polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G.

[1960] SEQ ID NO:1771 (encoded by SEQ ID NO: 11771) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion SV40 NLS, UL12 polypeptide, CTIP polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G.

[1961] SEQ ID NO:1772 (encoded by SEQ ID NO: 11772) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion SV40 NLS, STO7D polypeptide, CTIP polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G.

[1962] SEQ ID NO:1773 (encoded by SEQ ID NO: 11773) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of paRecT polypeptide and further includes amino acid substitutions D1133P, R669L, D537Y, T1011A, F148K, F973E, K1085R, K547R, N363H, A1282G, G257R, A442P, E399M, V539G.

[1963] SEQ ID NO:1774 (encoded by SEQ ID NO: 11774) sets forth the amino acid sequence of a R2-TG polypeptide having an N-terminal truncation (up to residue V161) with N-terminal fusion of paRecT polypeptide and further includes amino acid substitutions D555K, P280K, D1133P, R669L, D537Y, T1011A, F148K, F973E, K1085R, K547R.

[1964] SEQ ID NO:1775 (encoded by SEQ ID NO: 11775) sets forth the amino acid sequence of a R2-TG including amino acid substitutions C145G, F148K, E214R, P280K, E399M, D537Y, D555K, K700D, S895K, D923K, T989L, T1011S, I1219N, Y1242K, A1282G.

[1965] SEQ ID NO:1776 (encoded by SEQ ID NO: 11776) sets forth the amino acid sequence of a R2-TG including amino acid substitutions P280K, F148K, D555K, K700D, F667G, Y1242K, T989L, C145G, E542K, N889G, D923K, C518Y, E630D, L316R, E400M, Y571H, D1075E.

[1966] SEQ ID NO:1777 (encoded by SEQ ID NO: 11777) sets forth the amino acid sequence of a R2-TG including amino acid substitutions G257R, D555K, C145G, E214R, L213K, D537Y, A1282G, P280K, F148K, K700D, F667G, Y1242K, T989L, E542K, N889G, D923K, C518Y.

[1967] SEQ ID NO:1778 (encoded by SEQ ID NO: 11778) sets forth the amino acid sequence of a R2-TG including amino acid substitutions D1133P, R669L, D537Y, T1011A, F148K, F973E, K1085R, K547RR, N363H, A1282G, G257R, A442P, E399M.

[1968] SEQ ID NO:1779 (encoded by SEQ ID NO: 11779) sets forth the amino acid sequence of a R2-TG chimera. The RT domain of R2-TG is replaced with the Rt RTX and the RT-TG further includes amino acid substitutions D555K, P280K, F148K, T1011S, I1219N.

[1969] SEQ ID NO:1780 (encoded by SEQ ID NO: 11780) sets forth the amino acid sequence of a R2-TG chimera with N-terminal truncation (up to residue V161) with N-terminal fusion SV40 NLS, CTIP protein, the RT domain of R2-TG is replaced with the Rt RTX and the RT-TG further includes amino acid substitutions D555K, P280K, T1011S, I1219N.

[1970] FIGS. 70A and 70B is a bar graph showing results for drivers with the GFP reporter construct having a nucleic acid sequence set forth in SEQ ID NO: 175.

[1971] The data set included the following controls: Lipofectamine™-only and WT R2-TG GFP reporter SEQ ID NO:175 with R2-TG dead driver (SEQ ID NO: 1390 + SEQ ID NO: 175) as negative controls; WT RT-TG driver with WT reporter control (SEQ ID NO:1 +SEQ ID NO:175) and engineered RT-TG driver with WT reporter control (SEQ ID NO:636 +SEQ ID NO:175) as positive controls and a positive control for high integration SEQ ID NO:1716.

[1972] Several engineered R2-TG drivers exhibited improved activity above that of wild-type R2-TG: SEQ ID NO:1769, SEQ ID NO:1770, SEQ ID NO:1771, SEQ ID NO:1772, SEQ ID NO:1775, SEQ ID NO:1776 and SEQ ID NO:1777. In particular, SEQ ID NO:1770 which includes N-terminal truncation up to residue V161, N-terminal fusion SV40 NLS, BRCA2 polypeptide and DNA repair recruitment peptide CTIP and the amino acid substitutions P280K, D555K, T1011S, I1219N. However, none of the modification shows better activity than the engineered R2-TG SEQ ID NO:636.

Example 71: Evaluation of new R2 PROTEINS

[1973] Heterologous wild-type R2 proteins of R2 retroelements from different organisms were identified by scanning for insertions of retrotransposable elements in genomic rDNA genes of the different organisms available in public databases. Each identified R2 retroelement was further analyzed to determine the amino acid sequence of the wild-type (WT) R2 protein encoded by the open reading frame (ORF), as well as to identify the nucleotide sequence of the 5' retroelement UTR (5'RTE-UTR) and the 3' retroelement UTR (3'RTE-UTR) flanking the ORF. The newly identified R2 proteins are listed in Table 2, supra. Each entry provides the name of the species name of the organism from which the R2 retroelement was identified, the amino acid sequence of the wild-type R2 protein encoded by the ORF in the retroelement, and the nucleotide sequences of the 5' RTE-UTR and the 3' RTE-UTR.

[1974] An integration assay was performed to evaluate the effectiveness of the newly identified wild-type R2 proteins in integrating transgenes into genomic DNA. The assay was performed broadly as described in Example 2 which some changes as specified herein below. The effectiveness of a given R2 protein was indicated by the percentage of cells in which a GFP transgene comprised in an R2 reporter construct was successfully integrated by transgene integration by the R2 reporter.

[1975] The cells used in the study were U2OS cells (human osteosarcoma cell line) plated at 100K cells/well. Each well was treated with two RNA constructs in trans: a) a driver construct encoding a wild type R2 protein; and b) a GFP reporter construct. The GFP reporter construct comprised an antisense expression cassette for GFP (driven by a myeloproliferative sarcoma virus (MND) promoter and containing a polyadenylation signal from the thymidine kinase gene from HSV1), the RTE5'UTR and RTE3'UTR from the same species of organism from which

the R2 protein was derived, and a pair of 20-nucleotide homology arms to the genomic 28S rDNA integration site. The driver construct comprised the 5' RTE-UTR, a single ORF encoded by the given R2 retroelement and a polyadenylated SV40 sequence ending in A30-N10-A70 (wherein A30 is a sequence of 30 adenosine residues, N10 is a non-polyA sequence, and A70 is a sequence of additional 70 adenosine residues). The RNA constructs were produced by IVT as described above in Example 2.

[1976] U2OS cells were plated at 100K cells/well. RNA was transfected using Lipofectamine™. Integration was assessed based on the percentage of GFP positive cells (% GFP positive cells) after 24h, with a higher percentage of GFP positive cells being indicative of higher levels of integration. % GFP positive cells was assessed by FACS 24 hours after transfection with RNA. Results shown here are an average of at least two biological replicates.

[1977] FIGS. 71A and 71B illustrate the results of integration assays using different R2 retrotransposable elements in U2OS cells. The different retrotransposable element constructs were used in the trans configuration (driver and GFP reporter).

Example 72: Domain additions and amino acid substitutions improve R2 systems.

[1978] Driver constructs encoding engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOS: 636, 1799 and 1822-1834 described, hereinbelow were generated and tested with the GFP reporter constructs disclosed hereinbelow.

[1979] SEQ ID NO:1822 (encoded by SEQ ID NO: 11822) sets forth the amino acid sequence of a R2-TG polypeptide with amino acid substitutions P280K, F148K, D555K, K700D, F667G, Y1242K, T989L, C145G, E542K, N889G, D923K, C518Y, E630D, L316R, E400M, Y571H, D1075E, M671A, E214R, L213K, T1016L and E350R.

[1980] SEQ ID NO:1823 (encoded by SEQ ID NO: 11823) sets forth the amino acid sequence of a R2-TG polypeptide with amino acid substitutions G257R, D555K, C145G, E214R, L213K, D537Y, A1282G, P280K, F148K, K700D, F667G, Y1242K, T989L, E542K, N889G, D923K, C518Y, E630D, L316R, E400M.

[1981] SEQ ID NO:1824 (encoded by SEQ ID NO: 11824) sets forth the amino acid sequence of a R2-TG polypeptide with amino acid substitutions D1133P, R669L, D537Y, T1011A,

F148K, F973E, K1085R, K547R, N363H, A1282G, G257R, A442P, E399M, V539G, R1177G, H423Q and D555K.

[1982] SEQ ID NO:1825 (encoded by SEQ ID NO: 11825) sets forth the amino acid sequence of a R2-TG polypeptide with amino acid substitutions D555K, P280K, D1133P, R669L, D537Y, T1011A, F148K, F973E, K1085R, K547R, N363H, A1282G, G257R, A442P, E399M, V539G, R1177G, H423Q.

[1983] SEQ ID NO:1826 (encoded by SEQ ID NO: 11826) sets forth the amino acid sequence of a R2-TG polypeptide with N-terminal fusion of paRecT and amino acid substitutions D555K P280K D1133P R669L D537Y T1011A F148K F973E K1085R K547R N363H A1282G G257R A442P E399M.

[1984] SEQ ID NO:1827 (encoded by SEQ ID NO: 11827) sets forth the amino acid sequence of a R2-TG polypeptide with amino acid substitutions F148K P280K D555K T1011S I1219N A1282G G257R.

[1985] SEQ ID NO:1828 (encoded by SEQ ID NO: 11828) sets forth the amino acid sequence of a R2-TG polypeptide with N-terminal truncation of 161 residues, N-terminal fusion of SV40 NLS, CTIP, and Sto7D, and amino acid substitutions P280K D555K T1011S I1219N A1282G K700D C518Y.

[1986] SEQ ID NO:1829 (encoded by SEQ ID NO: 11829) sets forth the amino acid sequence of a R2-TG polypeptide with amino acid substitutions sets forth the amino acid sequence of a R2-TG polypeptide with N-terminal truncation of 161 residues, N-terminal fusion of SV40 NLS, CTIP, and Sto7D, and amino acid substitutions P280K D555K T1011S I1219N A1282G K700D C518Y T989L.

[1987] SEQ ID NO:1830 (encoded by SEQ ID NO: 11830) sets forth the amino acid sequence of a R2-TG polypeptide with N-terminal truncation of 161 residues, N-terminal fusion of SV40 NLS, VPX, CTIP, HMG, and amino acid substitutions P280K D555K T1011S I1219N A1282G.

[1988] SEQ ID NO:1831 (encoded by SEQ ID NO: 11831) sets forth the amino acid sequence of a R2-TG polypeptide with N-terminal truncation (up to residue V161) with N-terminal fusion

of a SV40 NLS, TORi peptide, CTIP polypeptide, HMGN1 polypeptide. R2-TG polypeptide includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G.

[1989] SEQ ID NO:1832 (encoded by SEQ ID NO: 11832) sets forth the amino acid sequence of a R2-TG polypeptide with N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, I53 peptide, CTIP polypeptide, HMGN1 polypeptide. R2-TG polypeptide includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G.

[1990] SEQ ID NO:1833 (encoded by SEQ ID NO: 11833) sets forth the amino acid sequence of a R2-TG polypeptide with N-terminal truncation of 161 residues, N-terminal fusion of SV40 NLS, CTIP, HMG, C-terminal fusion of VPX and amino acid substitutions P280K D555K T1011S I1219N A1282G.

[1991] SEQ ID NO:1834 (encoded by SEQ ID NO: 11834) sets forth the amino acid sequence of a R2-TG polypeptide with N-terminal truncation of 161 residues, N-terminal fusion of SV40 NLS, CTIP, HMG, C-terminal fusion of TORi and amino acid substitutions P280K D555K T1011S I1219N A1282G.

[1992] SEQ ID NO:1799 (encoded by SEQ ID NO: 11799) sets forth the amino acid sequence of a R2-TG polypeptide with N-terminal truncation of 161 residues, N-terminal fusion of SV40 NLS, CTIP, HMG, C-terminal fusion of RecT and amino acid substitutions P280K D555K T1011S I1219N A1282G.

[1993] SEQ ID NO:636 (encoded by SEQ ID NO: 10636) sets forth the amino acid sequence of a R2-TG polypeptide with N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G.

[1994] FIGS. 72A and 72B are bar graphs showing results of integration assays using engineered R2-TG driver having amino acid sequences set forth in SEQ ID NOS: 636, 1799 and 1822-1834. In this experiment, different retrotransposable element driver constructs were used with their respective R2 GFP reporter construct. Transfection was carried out with driver and reporter mRNA delivered with Lipofectamine™ to U2OS and full-length integration assessed by FACS GFP 24h. The drivers were engineered R2 protein constructs described above. The data set include the following controls: Non-treated control, R2-TG GFP reporter with WT R2-TG

control (SEQ ID NO: 10001 + SEQ ID NO: 175) and the engineered R2-TG positive control (SEQ ID NO: 10636+ SEQ ID NO: 175) as positive controls. Error bars indicate standard error of two replicates.

[1995] Many of the engineered constructs had higher activity than the WT R2-TG system, including: SEQ ID NO:1822, SEQ ID NO:1823, SEQ ID NO:1827, SEQ ID NO:1828, SEQ ID NO:1829, SEQ ID NO:1830, SEQ ID NO:1831, SEQ ID NO:1832, SEQ ID NO:1834, SEQ ID NO:636. SEQ ID NO:636 had the highest activity in this set.

Example 73: Domain additions and amino acid substitutions improve R2 systems.

[1996] Driver constructs encoding engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOS: 251, 620-622, 636, 1798 and 1821 described, hereinbelow were generated and tested with the GFP reporter construct having a nucleic acid sequence set forth in SEQ ID NO: 175.

[1997] SEQ ID NO:251 (encoded by SEQ ID NO: 10251) sets forth the amino acid sequence of a R2-TG polypeptide with N-terminal truncation of 161 residues and fusion of SV40NLS and Sto7D.

[1998] SEQ ID NO:620 (encoded by SEQ ID NO: 10620) sets forth the amino acid sequence of a R2-TG polypeptide with N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, Sto7D polypeptide, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, and A1282G.

[1999] SEQ ID NO:621 (encoded by SEQ ID NO: 10621) sets forth the amino acid sequence of a R2-TG polypeptide with N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, Sto7D polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G

[2000] SEQ ID NO:622 (encoded by SEQ ID NO: 10622) sets forth the amino acid sequence of a R2-TG polypeptide with N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide, Sto7D polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G

[2001] SEQ ID NO:1821 (encoded by SEQ ID NO: 11821) sets forth the amino acid sequence of a R2-TG polypeptide with N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and C-terminal fusion of I53 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G.

[2002] SEQ ID NO:1798 (encoded by SEQ ID NO: 11798) sets forth the amino acid sequence of a R2-TG polypeptide with N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G. The DNA sequence encoding the R2-TG reverse transcriptase is codon optimized for high frequency codons.

[2003] SEQ ID NO:636 (encoded by SEQ ID NO: 10636) is described herein, supra.

[2004] Reporter constructs are disclosed hereinbelow:

[2005] SEQ ID NO:1747 is a R2-TG GFP reporter with a 4 base-pair homology arm and 20bp polyA tail.

[2006] SEQ ID NO:1748 is a R2-TG GFP reporter with a 20 base-pair homology arm and 20bp polyA tail.

[2007] SEQ ID NO: 1749 is a R2-TG GFP reporter with a tev 5'UTR fused 5' of the 20bp homology arm.

[2008] FIGS. 73A and 73B are bar graphs showing results of integration assays using engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOs: 251, 620-622, 636, 1798 and 1821. In this experiment, different retrotransposable element driver constructs were used with their respective R2 GFP reporter construct. Transfection was carried out with driver and reporter mRNA delivered with Lipofectamine™ to U2OS cells and full-length integration assessed by FACS GFP 24h. The drivers were engineered R2 protein constructs described above. The data set included the following controls: Non-treated control, R2-TG GFP reporter SEQ ID NO:175 negative control, and WT R2-TG control (SEQ ID NO: 10001 + SEQ ID NO: 175) and an additional engineered R2-TG control (SEQ ID NO: 10636+ SEQ ID NO: 175) as positive controls. Error bars indicate standard error of two replicates.

[2009] Many of the engineered constructs had higher activity than the WT R2-TG system, including: SEQ ID NO:620, SEQ ID NO:621, SEQ ID NO:622 (encoded by SEQ ID NO: 10622), SEQ ID NO:1798, SEQ ID NO:636. In these experimental conditions, the fusion of SV40 NLS, CTIP, and HMGN as present in SEQ ID NO:636 had the highest activity, although fusions of SV40 NLS, Sto7D, CTIP, and HMGN in different orders gave enhanced activity as well.

[2010] The R2-TG GFP reporter SEQ ID NO:1749 which has a TEV 5'UTR fused upstream of the 5' homology arm had enhanced activity with both SEQ ID NO:636 and WT R2-TG, indicating potential stability improvements of the RNA template by this modification.

Example 74: Domain additions and amino acid substitutions improve R2 systems.

[2011] Driver constructs encoding engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOS: 629, 636, 1599, 1769, 1795-1798, 1800-1805 described, hereinbelow were generated and tested with the GFP reporter construct having a nucleic acid sequence set forth in SEQ ID NO: 175.

[2012] SEQ ID NO:1599 (encoded by SEQ ID NO: 11599) sets forth the amino acid sequence of a R2-TG polypeptide with N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide. R2-TG polypeptide includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G. The DNA sequence encoding the R2-TG reverse transcriptase is codon optimized for high frequency codons.

[2013] SEQ ID NO:1769 (encoded by SEQ ID NO: 11769) sets forth the amino acid sequence of a R2-TG polypeptide with amino acid substitutions P280K, D555K, T1011S, I1219N. The DNA sequence is codon optimized for high frequency codons.

[2014] SEQ ID NO:1795 (encoded by SEQ ID NO: 11795) sets forth the amino acid sequence of a R2-TG polypeptide with N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, and C-terminal fusion of HMGN1. R2-TG polypeptide includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G. The DNA sequence encoding the R2-TG reverse transcriptase is codon optimized for high frequency codons.

[2015] SEQ ID NO:1796 (encoded by SEQ ID NO: 11796) sets forth the amino acid sequence of a R2-TG polypeptide with N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, Dengue virus RNA helicase polypeptide, CTIP polypeptide. R2-TG polypeptide includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G. The DNA sequence encoding the R2-TG reverse transcriptase is codon optimized for high frequency codons.

[2016] SEQ ID NO:1797 (encoded by SEQ ID NO: 11797) sets forth the amino acid sequence of a R2-TG polypeptide with N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, C-terminal Dengue virus RNA helicase polypeptide. R2-TG polypeptide includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G. The DNA sequence encoding the R2-TG reverse transcriptase is codon optimized for high frequency codons.

[2017] SEQ ID NO:1798 (encoded by SEQ ID NO: 11798) sets forth the amino acid sequence of a R2-TG polypeptide with N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, HMGN1 polypeptide, CTIP polypeptide. R2-TG polypeptide includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G.

[2018] SEQ ID NO:1800 (encoded by SEQ ID NO: 11800) sets forth the amino acid sequence of a R2-TG polypeptide with C-terminal fusion of HCV peptide, and amino acid substitutions P280K, D555K, T1011S, I1219N. The DNA sequence is codon optimized for high frequency codons.

[2019] SEQ ID NO:1801 (encoded by SEQ ID NO: 11801) sets forth the amino acid sequence of a R2-TG polypeptide with C-terminal fusion of JFNS3 peptide and amino acid substitutions P280K, D555K, T1011S, I1219N. The DNA sequence is codon optimized for high frequency codons.

[2020] SEQ ID NO:1802 (encoded by SEQ ID NO: 11802) sets forth the amino acid sequence of a R2-TG polypeptide with C-terminal fusion of EshA peptide and amino acid substitutions P280K, D555K, T1011S, I1219N. The DNA sequence is codon optimized for high frequency codons.

[2021] SEQ ID NO:1803 (encoded by SEQ ID NO: 11803) sets forth the amino acid sequence a R2-AN polypeptide.

[2022] SEQ ID NO:1804 (encoded by SEQ ID NO: 11804) sets forth the DNA sequence of R2-TG GFP reporter with human globin beta 5' UTR appended upstream of homology arm.

[2023] SEQ ID NO:1805 (encoded by SEQ ID NO: 11805) sets forth the DNA sequence of R2-TG GFP reporter TMV 5' UTR appended upstream of homology arm.

[2024] SEQ ID NO:629 (encoded by SEQ ID NO: 10629) sets forth the amino acid sequence of a R2-TG polypeptide with N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, Sso7D polypeptide, CTIP polypeptide. R2-TG polypeptide includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G. The DNA sequence encoding the R2-TG reverse transcriptase is codon optimized for high frequency codons.

[2025] SEQ ID NO:636 (encoded by SEQ ID NO: 10636) is described herein supra.

[2026] FIG. 74 is a bar graph showing results of integration assays using engineered R2-TG drivers amino acid substitution having amino acid sequences set forth in SEQ ID NOS: 629, 636, 1599, 1769, 1795-1798, 1800-1805. In this experiment, different retrotransposable element driver constructs were used with their respective R2 GFP reporter construct. Transfection was carried out with driver and reporter mRNA delivered with Lipofectamine™ to U2OS and full-length integration assessed by FACS GFP 24h. The drivers were engineered R2 protein constructs described above. The data set included the following controls: Non-treated control, R2-TG GFP reporter SEQ ID NO:175 negative control, WT R2-TG control (SEQ ID NO: 10001 + SEQ ID NO: 175) positive control, additional engineered R2-TG positive control (SEQ ID NO: 10636+ SEQ ID NO: 175). Error bars indicate standard error of two replicates.

[2027] Many of the engineered constructs had higher activity than the WT R2-TG control, including SEQ ID NO:1599, SEQ ID NO:629, SEQ ID NO:1795, SEQ ID NO:1796, SEQ ID NO:1798, SEQ ID NO:1769, SEQ ID NO:1800, SEQ ID NO:636, SEQ ID NO:1804. In this experiment, SEQ ID NO:1796 had the highest activity. Compared to SEQ ID NO:636, it has addition of N-terminal Dengue virus RNA helicase followed by CTIP polypeptide. In comparison, SEQ ID NO:1797 with C-terminal fusion of Dengue virus RNA helicase had reduced activity. SEQ ID NO:1798 has the same amino acid substitutions and domain fusions as

SEQ ID NO:636 but the HMGN1 precedes the CTIP domain- a slight decrease in activity of SEQ ID NO:1798 relative to SEQ ID NO:636 indicates that CTIP preceding HMGN1 at the N-terminus of the R2-TG protein improves activity more. Similarly, comparing SEQ ID NO:1795 to SEQ ID NO:636, SEQ ID NO:1795 has HMGN1 fused to the C-terminus instead of the N-terminus and has significantly decreased activity, indicating that HMGN1 improves activity when fused to the N-terminus.

Example 75: Domain additions and amino acid substitutions improve R2 systems.

[2028] Driver constructs encoding engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOS: 623, 624, 636, 1806-1820 described, hereinbelow were generated and tested with the GFP reporter construct having a nucleic acid sequence set forth in SEQ ID NO: 175.

[2029] SEQ ID NO:1806 (encoded by SEQ ID NO: 11806) sets forth the amino acid sequence of a R2-TG polypeptide with N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, D923K, C518Y.

[2030] SEQ ID NO:1807 (encoded by SEQ ID NO: 11807) sets forth the amino acid sequence of a R2-TG polypeptide with N-terminal truncation (up to residue E184) with N-terminal fusion of a SV40 NLS, Sto7D polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, N889G, C518Y.

[2031] SEQ ID NO:1808 (encoded by SEQ ID NO: 11808) sets forth the amino acid sequence of a R2-TG polypeptide with N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, Sto7D polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, N889G, C518Y.

[2032] SEQ ID NO:1809 (encoded by SEQ ID NO: 11809) sets forth the amino acid sequence of a R2-TG polypeptide with N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, HMGN polypeptide, Sto7D polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, V661K.

[2033] SEQ ID NO:1810 (encoded by SEQ ID NO: 11810) sets forth the amino acid sequence of a R2-TG polypeptide with N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, HMGN polypeptide, CTIP polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, K700D, C518Y.

[2034] SEQ ID NO:1811 (encoded by SEQ ID NO: 11811) sets forth the amino acid sequence of a R2-TG polypeptide with N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, T989L.

[2035] SEQ ID NO:1812 (encoded by SEQ ID NO: 11812) sets forth the amino acid sequence of a R2-TG polypeptide with N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, E399M.

[2036] SEQ ID NO:1813 (encoded by SEQ ID NO: 11813) sets forth the amino acid sequence of a R2-TG polypeptide with N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, K700D, C518Y, T989L.

[2037] SEQ ID NO:1814 (encoded by SEQ ID NO: 11814) sets forth the amino acid sequence of a R2-TG polypeptide with N-terminal fusion of paRecT polypeptide and further includes amino acid substitutions D555K P280K D1133P R669L D537Y T1011A F148K F973E K1085R K547R N363H A1282G G257R.

[2038] SEQ ID NO:1815 (encoded by SEQ ID NO: 11815) sets forth the amino acid sequence of a R2-TG polypeptide with N-terminal fusion of paRecT polypeptide and further includes amino acid substitutions D555K P280K D1133P R669L D537Y T1011A F148K F973E K1085R K547R N363H A1282G.

[2039] SEQ ID NO:1816 (encoded by SEQ ID NO: 11816) sets forth the amino acid sequence of a R2-TG polypeptide with N-terminal fusion of paRecT polypeptide. R2-TG polypeptide includes amino acid substitutions D1133P, R669L, D537Y, T1011A, F148K, F973E, K1085R, K547R, N363H, A1282G, G257R, A442P, E399M, V539G, R1177G, H423Q, D555K.

[2040] SEQ ID NO:1817 (encoded by SEQ ID NO: 11817) sets forth the amino acid sequence of a R2-TG polypeptide with amino acid substitutions D555K, P280K, D1133P, R669L, D537Y, T1011A, F148K, F973E, K1085R, K547R, N363H, A1282G, G257R, A442P, E399M, V539G.

[2041] SEQ ID NO:1818 (encoded by SEQ ID NO: 11818) sets forth the amino acid sequence of a R2-TG polypeptide with amino acid substitutions P280K, F148K, D555K, K700D, F667G, Y1242K, T989L, C145G, E542K, N889G, D923K, C518Y, E630D, L316R, E400M, Y571H, D1075E, M671A.

[2042] SEQ ID NO:1819 (encoded by SEQ ID NO: 11819) sets forth the amino acid sequence of a R2-TG polypeptide with amino acid substitutions G257R, D555K, C145G, E214R, L213K, D537Y, A1282G, P280K, F148K, K700D, F667G, Y1242K, T989L, E542K, N889G, D923K, C518Y, E630D.

[2043] SEQ ID NO:1820 (encoded by SEQ ID NO: 11820) sets forth the amino acid sequence of a R2-TG polypeptide with amino acid substitutions G257R D1133P, R669L, D537Y, T1011A, F148K, F973E, K1085R, K547RR, N363H, A1282G, G257R, A442P.

[2044] SEQ ID NO:623 (encoded by SEQ ID NO: 10623) sets forth the amino acid sequence of a R2-TG polypeptide with N-terminal truncation (up to residue E184) with N-terminal fusion of a SV40 NLS, Sto7D polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, K700D, C518Y, T989L.

[2045] SEQ ID NO:624 (encoded by SEQ ID NO: 10624) sets forth the amino acid sequence of a R2-TG polypeptide with N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, Sto7D polypeptide, HMGN1 polypeptide and further includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G, K700D, C518Y, T989L.

[2046] SEQ ID NO:636 (encoded by SEQ ID NO: 10636) is described herein, supra.

[2047] FIGS. 75A and 75B are bar graphs showing results of integration assays using engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOs: 623, 624, 636, 1806-1820. In this experiment, different retrotransposable element driver constructs were used with their respective R2 GFP reporter construct. Transfection was carried out with driver and reporter mRNA delivered with Lipofectamine™ to U2OS and full-length integration assessed by

FACS GFP 24h. The drivers were engineered R2 protein constructs described above. The data set included the following controls: Non-treated control, R2-TG GFP reporter SEQ ID NO:175 negative control, WT R2-TG control (SEQ ID NO: 10001 + SEQ ID NO: 175) and engineered R2-TG (SEQ ID NO: 10636+ SEQ ID NO: 175) as positive controls. Error bars indicate standard error of two replicates.

[2048] Many of the engineered constructs had higher activity than the WT R2-TG system control, including SEQ ID NO:1806, SEQ ID NO:623, SEQ ID NO:1807, SEQ ID NO:624, SEQ ID NO:1808, SEQ ID NO:1809, SEQ ID NO:1810 (, SEQ ID NO:1811, SEQ ID NO:1812, SEQ ID NO:1813, and SEQ ID NO:636. Compared to SEQ ID NO:636, SEQ ID NO:1806, SEQ ID NO:1811, SEQ ID NO:1812, and SEQ ID NO:1813 had higher activity in at least one experiment, which may indicate higher activity due to amino acid substitutions T989L, E399M, D923K+C518Y, and K700D+C518Y+T989L.

Example 76: Transferability of improving modifications to homologous R2 systems.

[2049] Driver constructs encoding engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOS: 1401-1410 described hereinbelow were generated and tested with the corresponding GFP reporter constructs having nucleic acid sequences set forth in any one of SEQ ID NOS: 175, and 1411-1415.

[2050] The effect of transferring the improving modifications from R2-TG to general R2 systems was investigated. R2 systems with engineered drivers derived from R2-XC (SEQ ID NO:30029), R2-LS (SEQ ID NO:30019), R2-MC (SEQ ID NO:30020), R2-GP (SEQ ID NO:30015), and R2-PTT (SEQ ID NO:30027) were modified with domain additions (SV40 NLS, CTIP, and HMGN), amino acid substitutions, and truncations inferred by alignment with an engineered R2-TG driver protein (SEQ ID NO:636). Various codon optimization strategies were also applied.

[2051] The R2 driver constructs are described hereinbelow:

[2052] SEQ ID NO:1401 sets forth the amino acid sequence of R2-PTT polypeptide N-terminal deletion of 130 residues, N-terminal fusion of SV40 NLS, CTIP, and HMGN.

[2053] SEQ ID NO:1402 sets forth the amino acid sequence of R2-LS polypeptide N-terminal deletion of 61 residues, N-terminal fusion of SV40 NLS, CTIP, and HMGN.

[2054] SEQ ID NO:1403 sets forth the amino acid sequence of R2-MC polypeptide N-terminal deletion of 64 residues, N-terminal fusion of SV40 NLS, CTIP, and HMGN.

[2055] SEQ ID NO:1404 sets forth the amino acid sequence of R2-GP polypeptide N-terminal deletion of 71 residues, N-terminal fusion of SV40 NLS, CTIP, and HMGN.

[2056] SEQ ID NO:1405 sets forth the amino acid sequence of R2-XC polypeptide N-terminal deletion of 71 residues, N-terminal fusion of SV40 NLS, CTIP, and HMGN.

[2057] SEQ ID NO:1406 sets forth the amino acid sequence of R2-PTT polypeptide N-terminal deletion of 130 residues, N-terminal fusion of SV40 NLS, CTIP, and HMGN, and amino acid substitutions P257K and N1326G.

[2058] SEQ ID NO:1407 sets forth the amino acid sequence of R2-LS polypeptide N-terminal deletion of 61 residues, N-terminal fusion of SV40 NLS, CTIP, and HMGN and amino acid substitutions P178K and N1132G.

[2059] SEQ ID NO:1408 sets forth the amino acid sequence of R2-MC polypeptide N-terminal deletion of 64 residues, N-terminal fusion of SV40 NLS, CTIP, and HMGN, and amino acid substitutions P191K and A877S.

[2060] SEQ ID NO:1409 sets forth the amino acid sequence of R2-GP polypeptide N-terminal deletion of 71 residues, N-terminal fusion of SV40 NLS, CTIP, and HMGN, and amino acid substitutions P178K, I1068N and N1131G.

[2061] SEQ ID NO:1410 sets forth the amino acid sequence of R2-XC polypeptide N-terminal deletion of 71 residues, N-terminal fusion of SV40 NLS, CTIP, and HMGN, and amino acid substitutions P179K and N1138G.

[2062] The corresponding R2 GFP reporter constructs are:

[2063] SEQ ID NO:1411 sets forth the DNA sequence of R2-PTT GFP reporter.

[2064] SEQ ID NO:1412 sets forth the DNA sequence of R2-LS GFP reporter.

[2065] SEQ ID NO:1413 sets forth the DNA sequence of R2-MC GFP reporter.

[2066] SEQ ID NO:1414 sets forth the DNA sequence of R2-GP GFP reporter.

[2067] SEQ ID NO:1415 sets forth the DNA sequence of R2-XC GFP reporter.

[2068] FIG. 76A is a bar graph showing results of integration assays using engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOs: 1401, 1406, . In this experiment, different retrotransposable element driver constructs were used with their respective R2 GFP reporter construct. Transfection was carried out with driver and reporter mRNA delivered with Lipofectamine™ to HUH7 and full-length integration assessed by dPCR targeting transgene polyadenylation signal at 24h. The drivers were engineered R2 protein constructs described above. The data set included the following controls: Non-treated control, WT R2-TG GFP reporter SEQ ID NO:175 and R2-PTT GFP reporter SEQ ID NO:1411 as negative controls; WT R2-TG control (SEQ ID NO: 10001 + SEQ ID NO: 175) as positive control. Error bars indicate standard error of two replicates.

[2069] According to dPCR to quantify full-length integration, modification of R2-PTT driver protein by truncation of 130 residues and addition of N-terminal SV40 NLS, CTIP, and HMGN can improve activity greater than 4-fold, while addition of the P257K and N1326G amino acid substitutions did not improve activity. These results indicate that domain fusions from R2-TG can be transferred to homologous R2-PTT.

[2070] FIG. 76B is a bar graph showing results of integration assays engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOs: 1402, 1405, 1407, 1410. In this experiment, different retrotransposable element driver constructs were used with their respective R2 GFP reporter construct. Transfection was carried out with driver and reporter mRNA delivered with Lipofectamine™ to HUH7 and full-length integration assessed by dPCR targeting transgene polyadenylation signal at 24h. The drivers were engineered R2 protein constructs described above. The data set included the following controls: Non-treated control, R2-TG GFP reporter SEQ ID NO:175 and R2-LS GFP reporter SEQ ID NO:1412 and R2-XC GFP reporter SEQ ID NO:1415 as negative controls, WT R2-TG control (SEQ ID NO: 10001 + SEQ ID NO: 175) as positive control. Error bars indicate standard error of two replicates.

[2071] According to dPCR to quantify full-length integration, modification of R2-LS driver protein by truncation of 61 residues and addition of N-terminal SV40 NLS, CTIP, and HMGN can improve activity greater than 5-fold, and addition of P178K and N1132G amino acid substitutions also improved activity on top of this. Modification of R2-XC driver protein by truncation of 71 residues and addition of N-terminal SV40 NLS, CTIP, and HMGN can improve activity greater than 5-fold, and addition of P179K and N1138G amino acid substitutions further improved activity. These results indicate that domain fusions and amino acid substitutions can improve R2-LS and R2-XC systems.

[2072] FIG. 76C is a bar graph showing results of integration assays using engineered R2-TG having amino acid sequences set forth in SEQ ID NOs: 1403, 1404. In this experiment, different retrotransposable element driver constructs were used with their respective R2 GFP reporter construct. Transfection was carried out with driver and reporter mRNA delivered with Lipofectamine™ to HUH7 and full-length integration assessed by dPCR targeting transgene polyadenylation signal at 24h. The drivers were engineered R2 protein constructs described above. The data set included the following controls: Non-treated control, R2-TG GFP reporter SEQ ID NO:175, R2-GP GFP reporter SEQ ID NO:1413, and R2-MC GFP reporter SEQ ID NO:1414 as negative controls, WT R2-TG control (SEQ ID NO: 10001 + SEQ ID NO: 175) as positive control. Error bars indicate standard error of two replicates.

[2073] According to dPCR to quantify full-length integration, modification of R2-GP driver protein by truncation of 71 residues and addition of N-terminal SV40 NLS, CTIP, and HMGN can improve activity greater than 2-fold. Modification of R2-MC driver protein by truncation of 64 residues and addition of N-terminal SV40 NLS, CTIP, and HMGN can improve activity approximately 50%. These results indicate that domain fusions can improve activity of R2-GP and R2-MC systems.

[2074] These results indicate that the modifications that improve R2-TG can improve other R2 systems in cell line HUH7.

[2075] To test whether engineered R2 systems can improve activity in human tissue, engineered constructs were tested in primary human hepatocytes with LNP-mRNA delivery.

[2076] FIG. 76D is a bar graph showing results of integration assays using engineered R2-TG having amino acid sequences set forth in SEQ ID NOs: 1417-1421. In this experiment, different retrotransposable element driver constructs were used with their respective R2 GFP reporter construct. Transfection was carried out with driver and reporter mRNA delivered with LNP+mRNA to primary human hepatocytes and full-length integration assessed by dPCR targeting transgene polyadenylation signal at 24h. The drivers were engineered R2 protein constructs described above. The data set included the following controls: Non-treated control, R2-TG GFP reporter SEQ ID NO:175, R2-PTT GFP reporter SEQ ID NO:1411, R2-LS GFP reporter SEQ ID NO:1412 and ancestral R2 GFP reporter SEQ ID NO:581 as negative controls, WT R2-TG control (SEQ ID NO: 10001 + SEQ ID NO: 175) as positive control. Error bars indicate standard error of two replicates.

[2077] SEQ ID NO:1417 sets forth the DNA sequence of R2-PTT polypeptide with 5' TEV UTR and 3' HAG UTR.

[2078] SEQ ID NO:1418 sets forth the DNA sequence of R2-PTT polypeptide with 5' TEV UTR and 3' HAG UTR. The sequence encoding R2-PTT is codon optimized for high frequency codons.

[2079] SEQ ID NO:1419 sets forth the DNA sequence of R2-PTT polypeptide with 5' TEV UTR and 3' HAG UTR. The sequence encoding R2-PTT is codon optimized for fast translation codons.

[2080] SEQ ID NO:1420 sets forth the DNA sequence of R2-LS polypeptide with 5' TEV UTR and 3' HAG UTR.

[2081] SEQ ID NO:1421 sets forth the DNA sequence of R2-LS polypeptide with 5' TEV UTR and 3' HAG UTR. The sequence encoding R2-PTT is codon optimized for fast translation codons.

[2082] According to dPCR to quantify full-length integration, modification of R2 drivers through codon optimization and domain additions can improve integration in a human tissue such as primary human hepatocytes to deliver full-length transgenes.

[2083] Notably, the combination of an engineered R2-TG (SEQ ID NO: 10636) with R2-PTT reporter template (SEQ ID NO:1411) had the highest integration in this set.

Example 77: Engineered R2 systems have improved activity in primary human tissue when delivered as LNP-mRNA.

[2084] R2 systems engineered by domain additions and amino acid substitutions were tested for integration of full-length transgene in human tissue (primary human hepatocytes) when delivered as LNP-mRNA. Top constructs from cell line screening (U2OS and HUH7 cell lines) were tested in primary human hepatocytes for functional integration of GFP transgene.

[2085] **Ex. 77A.** The driver constructs are described, hereinbelow:

[2086] SEQ ID NO:1368 (encoded by SEQ ID NO: 11368) sets forth the amino acid sequence of a R2-TG polypeptide with amino acid substitutions P280K, D555K, P739R, T1011S, I1219N and 3'UTR HAG (SEQ ID NO: 3025).

[2087] SEQ ID NO:1369 (encoded by SEQ ID NO: 11369) sets forth the amino acid sequence of a R2-TG polypeptide with amino acid substitutions P280K, D923K, D555K and 3'UTR HAG (SEQ ID NO: 3025).

[2088] SEQ ID NO:1370 (encoded by SEQ ID NO: 11370) sets forth the amino acid sequence of a R2-TG polypeptide with amino acid substitutions K700D and P280K and 3'UTR HAG (SEQ ID NO: 3025).

[2089] SEQ ID NO:1371 (encoded by SEQ ID NO: 11371) sets forth the amino acid sequence of a R2-TG polypeptide with amino acid substitutions D555K, K700D, P280K, C518Y and 3'UTR HAG (SEQ ID NO: 3025).

[2090] SEQ ID NO:1372 (encoded by SEQ ID NO: 11372) sets forth the amino acid sequence of a R2-TG polypeptide with amino acid substitutions P280K, Y1242K and 3'UTR HAG (SEQ ID NO: 3025).

[2091] SEQ ID NO:1374 (encoded by SEQ ID NO: 11374) sets forth the amino acid sequence of a R2-TG polypeptide with 3'UTR HAG (SEQ ID NO: 3025).

[2092] SEQ ID NO:1375 (encoded by SEQ ID NO: 11375) sets forth the amino acid sequence of a R2-TG polypeptide with amino acid substitutions with D555K amino acid substitution and 3'UTR HAG (SEQ ID NO: 3025).

[2093] Driver constructs encoding engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOS: 1368-1372 and 1374-1375 described, hereinbelow were generated and tested with the GFP reporter construct having a nucleic acid sequence set forth in SEQ ID NO: 175. **FIG. 77A** is a bar graph showing results of integration assays using engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOS: 1368-1372 and 1374-1375. In this experiment, different retrotransposable element driver constructs were used with the R2 GFP reporter construct SEQ ID NO: 175. Transfection was carried out with driver and reporter mRNA delivered with LNP to primary human hepatocyte cells and full-length integration assessed by FACS at 72h. The drivers were engineered R2-TG protein constructs described above. The data set included the following controls: Non-treated control, GFP reporter SEQ ID NO:175 as negative control, WT RT-TG driver with WT reporter control (SEQ ID NO:1 +SEQ ID NO:175) as positive control. Error bars indicate standard error of two replicates.

[2094] The WT R2-TG control (SEQ ID NO:1 + SEQ ID NO:175) was found to give very low %GFP+ cells, <0.5%. Addition of HAG 3'UTR (mouse Hemoglobin Alpha 3'UTR (SEQ ID NO: 3025)) was able to raise the integration activity ~2-fold (SEQ ID NO:1374). Addition of combination of amino acid substitutions including P280K, D555K, D923K, K700D, P739R, T1011S, I1219N, Y1242K, K579R were able to raise the activity further. In particular, SEQ ID NO:1375 which contains the D555K amino acid substitution alone had a several fold increase in activity to ~2.5% GFP+ cells compared to the WT R2-TG control (SEQ ID NO:1 +SEQ ID NO:175).

[2095] **Ex. 77B:** The driver constructs having amino acid sequences set forth in SEQ ID NOS: 1376, 1378, 1380-1389 and 1396 were generated and tested with GFP reporter having a nucleic acid sequence set forth in SEQ ID NO: 175 and are described, hereinbelow:

[2096] SEQ ID NO:1376 (encoded by SEQ ID NO: 11376) sets forth the amino acid sequence of a R2-TG polypeptide with N-terminal fusion of SV40 NLS, amino acid substitutions F148K, P280K, D555K, T1011S, I1219N, A1282G. The encoding DNA contains 5'UTR from tobacco etch virus and 3'UTR HAG (SEQ ID NO: 3025).

[2097] SEQ ID NO:1378 (encoded by SEQ ID NO: 11378) sets forth the amino acid sequence of R2-TG polypeptide with amino acid substitutions F148K, P280K, D555K, T1011S, I1219N A1282G. The encoding DNA contains 5'UTR from tobacco etch virus and 3'UTR HAG (SEQ ID NO: 3025).

[2098] SEQ ID NO:1380 (encoded by SEQ ID NO: 11380) sets forth the amino acid sequence of R2-TG polypeptide with amino acid substitutions P280K, F148K, D555K. The encoding DNA contains 3'UTR HAG (SEQ ID NO: 3025).

[2099] SEQ ID NO:1381 (encoded by SEQ ID NO: 11381) sets forth the amino acid sequence of R2-TG polypeptide with amino acid substitutions P280K, F148K, D555K, K700D, F667G. The encoding DNA contains 3'UTR HAG (SEQ ID NO: 3025).

[2100] SEQ ID NO:1382 (encoded by SEQ ID NO: 11382) sets forth the amino acid sequence of R2-TG polypeptide with amino acid substitutions C145G, P280K, D555K, E638L, F667G, K700D, T1011S, I1219N, Y1242K. The encoding DNA contains 5'UTR from tobacco etch virus and 3'UTR HAG (SEQ ID NO: 3025).

[2101] SEQ ID NO:1383 (encoded by SEQ ID NO: 11383) sets forth the amino acid sequence of R2-TG polypeptide with amino acid substitutions D555K K700D P280K C518Y D923K. The encoding DNA contains 3'UTR HAG (SEQ ID NO: 3025).

[2102] SEQ ID NO:1384 (encoded by SEQ ID NO: 11384) sets forth the amino acid sequence of R2-TG polypeptide with amino acid substitutions D555K K700D P280K C518Y L213R. The encoding DNA contains 3'UTR HAG (SEQ ID NO: 3025).

[2103] SEQ ID NO:1385 (encoded by SEQ ID NO: 11385) sets forth the amino acid sequence of R2-TG polypeptide with amino acid substitutions D555K K700D P280K C518Y E350R. The encoding DNA contains 3'UTR HAG (SEQ ID NO: 3025).

[2104] SEQ ID NO:1386 (encoded by SEQ ID NO: 11386) sets forth the amino acid sequence of R2-TG polypeptide with amino acid substitutions G257R, D555K, C145G, E214R. The encoding DNA contains 3'UTR HAG (SEQ ID NO: 3025).

[2105] SEQ ID NO:1387 (encoded by SEQ ID NO: 11387) sets forth the amino acid sequence of R2-TG polypeptide with amino acid substitutions D555K K700D P280K C518Y D923K I1219N. The encoding DNA contains 3'UTR HAG (SEQ ID NO: 3025).

[2106] SEQ ID NO:1388 (encoded by SEQ ID NO: 11388) sets forth the amino acid sequence of R2-TG polypeptide N-terminal deletion of 161 residues, N-terminal fusion of SV40 NLS and CTIP, with amino acid substitutions P280K, D555K, P739R, T1011S, I1219N, A1282G. The encoding DNA contains 3'UTR HAG (SEQ ID NO: 3025).

[2107] SEQ ID NO:1389 (encoded by SEQ ID NO: 11389) sets forth the amino acid sequence of R2-TG polypeptide with N-terminal fusion of bipartite SV40 NLS and amino acid substitutions F148K, P280K, D555K, T1011S, I1219N, A1282G. The encoding DNA contains 3'UTR HAG (SEQ ID NO: 3025).

[2108] **FIG. 77B** is a bar graph showing results of integration assays using engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOS: 139, 139, 234, 258, 1376, 1378, 1380-1389 and 1396. In this experiment, different retrotransposable element driver constructs were used with the R2 GFP reporter construct SEQ ID NO: 175. Transfection was carried out with driver and reporter mRNA delivered with LNP to primary human hepatocyte cells and full-length integration assessed by FACS at 72h. The drivers were engineered R2-TG protein constructs described above. The data set included the following controls: Non-treated control, GFP reporter SEQ ID NO:175 negative control, WT RT-TG (SEQ ID NO: 1 + SEQ ID NO: 175) positive control. Error bars indicate standard error of two replicates.

[2109] Many of the constructs tested had higher integration efficiency than the WT R2-TG control (SEQ ID NO:1+SEQ ID NO:175): SEQ ID NO:1376, SEQ ID NO:1396, SEQ ID NO:1380, SEQ ID NO:1381, SEQ ID NO:1382, SEQ ID NO:1383, SEQ ID NO:1384, SEQ ID NO:1385, SEQ ID NO:1386, SEQ ID NO:1387, SEQ ID NO:1388.

[2110] Many of these constructs had a higher integration efficiency than SEQ ID NO:1375 (see FIG. 77A) which contained the D555K amino acid substitution at the protein level and the HAG 3'UTR at the DNA level. These constructs included: SEQ ID NO:1376, SEQ ID NO:139, SEQ ID NO:1380, SEQ ID NO:1381, SEQ ID NO:1382, SEQ ID NO:1383, SEQ ID NO:1384, SEQ ID NO:1386, SEQ ID NO:1387, SEQ ID NO:1388. In particular, the construct with the highest

activity SEQ ID NO:1380 with amino acid substitutions P280K, F148K, D555K which were present in most of the highest activity constructs tested.

[2111] **Ex.77C.** The driver constructs having amino acid sequences set forth in SEQ ID NOs: 1390, 1378, 1391, 1218, 636, 260, 618, 1392-1398, 1375, 1388, 1228 were generated and tested with GFP reporter having a nucleic acid sequence set forth in SEQ ID NO: 175 and are described, hereinbelow:

[2112] SEQ ID NO:1390 (encoded by SEQ ID NO: 11390) sets forth the amino acid sequence of R2-TG polypeptide with K1307A amino acid substitution.

[2113] SEQ ID NO:1378 (encoded by SEQ ID NO: 11378) sets forth the amino acid sequence of R2-TG polypeptide with amino acid substitutions F148K P280K D555K T1011S I1219N A1282G.

[2114] SEQ ID NO:1391 (encoded by SEQ ID NO: 11391) sets forth the amino acid sequence of R2-TG polypeptide N-terminal deletion of 161 residues, N-terminal fusion of SV40 NLS and CTIP, with amino acid substitutions P280K D555K T1011S I1219N A1282G.

[2115] SEQ ID NO:1218 (encoded by SEQ ID NO: 11218) sets forth the amino acid sequence of a R2-TG polypeptide with N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, HMGN1 polypeptide, CTIP polypeptide. R2-TG polypeptide includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G.

[2116] SEQ ID NO:636 (encoded by SEQ ID NO: 10636) sets forth the amino acid sequence of a R2-TG polypeptide with N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide. R2-TG polypeptide includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G.

[2117] SEQ ID NO:260 (encoded by SEQ ID NO: 10260) sets forth the amino acid sequence of a R2-TG polypeptide with N-terminal truncation (up to residue E184) with N-terminal fusion of a SV40 NLS and Sto7D.

[2118] SEQ ID NO:618 (encoded by SEQ ID NO: 10618) sets forth the amino acid sequence of a R2-TG polypeptide with N-terminal truncation (up to residue E184) with N-terminal fusion of a SV40 NLS and Sto7D and HMGN1.

[2119] SEQ ID NO:635 (encoded by SEQ ID NO: 10635) sets forth the amino acid sequence of a R2-TG polypeptide with N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS and Sto7D and CTIP.

[2120] SEQ ID NO:1392 (encoded by SEQ ID NO: 11392) sets forth the amino acid sequence of a R2-TG polypeptide with amino acid substitutions P280K D555K and reverse transcriptase domain replaced from homologous domain from R2-TOC.

[2121] SEQ ID NO:1393 (encoded by SEQ ID NO: 11393) sets forth the amino acid sequence of a R2-TG polypeptide with N-terminal truncation (up to residue E184) with N-terminal fusion of a SV40 NLS and HMGN1, amino acid substitutions P280K D555K and reverse transcriptase domain replaced from homologous domain from R2-TOC.

[2122] SEQ ID NO:1228 (encoded by SEQ ID NO: 11228) sets forth the amino acid sequence of a R2-TG polypeptide with N-terminal truncation (up to residue E184) with N-terminal fusion of a SV40 NLS and CTIP and HMGN1, amino acid substitutions P280K D555K and reverse transcriptase domain replaced from homologous domain from R2-TOC.

[2123] SEQ ID NO:1394 (encoded by SEQ ID NO: 11394) sets forth the amino acid sequence of a R2-TG polypeptide with N-terminal truncation (up to residue E184) with N-terminal fusion of a SV40 NLS and CTIP and HMGN1, C-terminal fusion of RNA binding domain, amino acid substitutions P280K, D555K, T1011S, I1219N.

[2124] SEQ ID NO:1395 (encoded by SEQ ID NO: 11395) sets forth the amino acid sequence of a R2-TG polypeptide with amino acid substitutions C145G, F148K, P280K, E399M, D555K, K700D, D923K, T989L, T1011S, I1219N, A1282G.

[2125] SEQ ID NO:1396 (encoded by SEQ ID NO: 11396) sets forth the amino acid sequence of R2-TG polypeptide with N-terminal fusion of RecT from *Pseudomonas aeruginosa* and amino acid substitutions D1133P, R669L, D537Y.

[2126] SEQ ID NO:1397 (encoded by SEQ ID NO: 11397) sets forth the amino acid sequence of a R2-TG polypeptide with N-terminal truncation (up to residue E184) with N-terminal fusion of SV40 NLS and HMGN1, C-terminal fusion of RNA binding domain, and amino acid substitutions P280K and D555K.

[2127] SEQ ID NO:1398 (encoded by SEQ ID NO: 11398) sets forth the amino acid sequence of a R2-TG polypeptide with N-terminal truncation (up to residue E184) with N-terminal fusion of a SV40 NLS and CTIP and HMGN1, C-terminal fusion of RNA binding domain, with amino acid substitutions P280K, D555K.

[2128] SEQ ID NO:1375 (encoded by SEQ ID NO: 11375) sets forth the amino acid sequence of a R2-TG polypeptide with D555K amino acid substitution and 3'UTR HAG (SEQ ID NO: 3025).

[2129] SEQ ID NO:1388 (encoded by SEQ ID NO: 11388) sets forth the amino acid sequence of R2-TG polypeptide N-terminal deletion of 161 residues, N-terminal fusion of SV40 NLS and CTIP, with amino acid substitutions P280K, D555K, P739R, T1011S, I1219N, A1282G. The encoding DNA contains 3'UTR HAG (SEQ ID NO: 3025).

[2130] FIG. 77C is a bar graph showing results of integration assays using engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOs: 1390, 1378, 1391, 1218, 636, 260, 618, 1392-1398, 1375, 1388, 1228. In this experiment, different retrotransposable element driver constructs were used with the R2 GFP reporter construct SEQ ID NO: 175. Transfection was carried out with driver and reporter mRNA delivered with LNP to primary human hepatocyte cells and full-length integration assessed by FACS at 72h. The drivers were engineered R2-TG protein constructs described above. The data set included the following controls: Non-treated control, GFP reporter SEQ ID NO:175 negative control, WT R2-TG control (SEQ ID NO:1 +SEQ ID NO:175) positive control. Error bars indicate standard error of two replicates.

[2131] Many of the constructs tested had higher integration efficiency than the WT R2-TG construct (SEQ ID NO:1), for example the R2-TG variant with activity-improving D555K amino acid substitution (SEQ ID NO:1375), and R2-TG variant with activity-improving amino acid substitutions F148K, P280K, D555K, T1011S, I1219N, A1282G (SEQ ID NO:1378); SEQ ID NO:1391, SEQ ID NO:1218, SEQ ID NO:636, SEQ ID NO:260, SEQ ID NO:618, and SEQ ID NO:635.

[2132] The construct with the highest activity is SEQ ID NO:618, which has an N-terminal fusion of SV40NLS, Sto7D, and HMGN1, as well as N-terminal deletion of 184 residues. In

comparison, SEQ ID NO:260 has an N-terminal fusion of SV40NLS, Sto7D, as well as N-terminal deletion of 184 residues, and approximately three-fold lower activity, showing the importance of HMGN1 addition. In turn, SEQ ID NO:260 has significantly higher activity than base R2-TG construct, indicating the activity-improvement from N-terminal fusion of Sto7D.

[2133] Comparing constructs SEQ ID NO:1391, SEQ ID NO:1218 and SEQ ID NO:636, which differ in inclusion of HMGN1 polypeptide, approximately two-fold activity improvement was observed by fusing HMGN1 polypeptide (SEQ ID NO:1391) vs SEQ ID NO:636 but only when it was C-terminal of the CTIP polypeptide (SEQ ID NO:1218 vs SEQ ID NO:636).

[2134] **Ex. 77D.** The driver constructs having amino acid sequences set forth in SEQ ID NOs: 636, 618, 783, and 1234 were generated and tested with GFP reporter having a nucleic acid sequence set forth in SEQ ID NO: 582 and are described, hereinbelow:

[2135] SEQ ID NO:636 (encoded by SEQ ID NO: 10636) sets forth the amino acid sequence of a R2-TG polypeptide with N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide. R2-TG polypeptide includes amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G.

[2136] SEQ ID NO:618 (encoded by SEQ ID NO: 10618) sets forth the amino acid sequence of a R2-TG polypeptide with N-terminal truncation (up to residue E184) with N-terminal fusion of a SV40 NLS and Sto7D and HMGN1.

[2137] SEQ ID NO:783 (encoded by SEQ ID NO: 10783) sets forth the amino acid sequence a R2-TG polypeptide with N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, EIF4A polypeptide, CTIP polypeptide, HMGN1 polypeptide and amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G.

[2138] SEQ ID NO:1234 (encoded by SEQ ID NO: 11234) sets forth the amino acid sequence a R2-TG polypeptide with N-terminal truncation (up to residue E184) with N-terminal fusion of a SV40 NLS, TUDOR polypeptide, CTIP polypeptide, HMGN1 polypeptide and amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G.

[2139] **FIG. 77D** is a graph showing results of integration assays using engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOs: 636, 618, 783, and 1234. In this

experiment, different retrotransposable element driver constructs were used with the R2 GFP reporter construct SEQ ID NO: 582. Transfection was carried out with driver and reporter mRNA delivered with LNP to hepatocarcinoma cell line HUH7 and primary human hepatocyte cells and full-length integration assessed by FACS and dPCR at 72h. The drivers were engineered R2-TG protein constructs described above. The data set included the following controls: Non-treated control, GFP reporter (SEQ ID NO:582) negative control. Error bars indicate standard error of two replicates.

[2140] SEQ ID NO:783 has improved activity relative to SEQ ID NO:636, which was previously shown in FIG. 77C to have higher activity than the WT R2-TG base construct (SEQ ID NO:1). This demonstrates that fusion of EIF4A human RNA helicase can improve integration activity.

[2141] SEQ ID NO:1234 has improved activity compared to SEQ ID NO:636, SEQ ID NO:783, and SEQ ID NO:618. It contains a TUDOR peptide derived from TDRD3 which recruits DHX9. This demonstrates improved integration activity by fusion of this peptide.

[2142] Ex. 77E. The driver constructs having amino acid sequences set forth in SEQ ID NOs: 636, 936, 1021, 618, 783, 1234-1236, 3028, 3029 were generated and tested with GFP reporter having a nucleic acid sequence set forth in SEQ ID NO: 175 and are described, hereinbelow:

[2143] SEQ ID NO:636 (encoded by SEQ ID NO: 10636) is described herein, supra.

[2144] SEQ ID NO:936 (encoded by SEQ ID NO: 10936) sets forth the amino acid sequence of a R2-TG polypeptide with N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, HMGN1 polypeptide. R2-TG polypeptide includes amino acid substitutions P280K, D555K, D923K, T1011S, A1038G, I1219N, A1282G.

[2145] SEQ ID NO:1021 (encoded by SEQ ID NO: 11021) sets forth the amino acid sequence of a R2-TG polypeptide with N-terminal truncation (up to residue E184) with N-terminal fusion of a SV40 NLS, CTIP polypeptide, and HMGN1 polypeptide. R2-TG polypeptide includes amino acid substitutions P280K, E399M, D555K, D923K, T989L, T1011S, R1177G, I1219N, A1282G.

[2146] SEQ ID NO:618 (encoded by SEQ ID NO: 10618) sets forth the amino acid sequence of a R2-TG polypeptide with N-terminal truncation (up to residue E184) with N-terminal fusion of a SV40 NLS and Sto7D and HMGN1.

[2147] SEQ ID NO:783 (encoded by SEQ ID NO: 10783) sets forth the amino acid sequence a R2-TG polypeptide with N-terminal truncation (up to residue V161) with N-terminal fusion of a SV40 NLS, EIF4A polypeptide, CTIP polypeptide, HMGN1 polypeptide and amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G.

[2148] SEQ ID NO:1234 (encoded by SEQ ID NO: 11234) sets forth the amino acid sequence a R2-TG polypeptide with N-terminal truncation (up to residue E184) with N-terminal fusion of a SV40 NLS, TUDOR polypeptide, Sto7D polypeptide, HMGN1 polypeptide and amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G.

[2149] SEQ ID NO:1235 (encoded by SEQ ID NO: 11235) sets forth the amino acid sequence a R2-TG polypeptide with N-terminal truncation (up to residue E184) with N-terminal fusion of a cMyc NLS, STO7D polypeptide, HMGN1 polypeptide and amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G.

[2150] SEQ ID NO:1236 (encoded by SEQ ID NO: 11236) sets forth the amino acid sequence a R2-TG polypeptide with N-terminal truncation (up to residue E184) with N-terminal fusion of a SV40 NLS, Sto7D polypeptide, HMGN1 polypeptide and amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G.

[2151] SEQ ID NO:3028 (encoded by SEQ ID NO: 13028) sets forth the amino acid sequence a R2-TG polypeptide with N-terminal truncation (up to residue E184) with N-terminal fusion of a SV40 NLS, Sto7D polypeptide, HMGN1 polypeptide and amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G. The DNA encoding this protein has codon optimization that optimizes for codon frequency.

[2152] SEQ ID NO:3029 (encoded by SEQ ID NO: 13029) sets forth the amino acid sequence a R2-TG polypeptide with N-terminal truncation (up to residue E184) with N-terminal fusion of a SV40 NLS, EIF4A polypeptide, Sto7D polypeptide, HMGN1 polypeptide and amino acid substitutions P280K, D555K, T1011S, I1219N, A1282G. The DNA encoding this protein has codon optimization that optimizes for codon frequency.

[2153] FIGS. 77E and 77F are graphs showing the results of integration assays using engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOs: 636, 936, 1021, 618, 783, 1234-1236, 3028, 3029. In this experiment, different retrotransposable element driver constructs were used with the R2 GFP reporter construct SEQ ID NO: 582. Transfection was carried out with driver and reporter mRNA delivered with LNP to primary human hepatocyte cells and full-length integration assessed by FACS and dPCR at 72h. The drivers were engineered R2-TG protein constructs described above. Error bars indicate standard error of two replicates.

[2154] The constructs tested here exhibited higher integration activity compared to engineered R2 construct SEQ ID NO:636, which was previously found to have much higher activity than the WT R2-TG control (SEQ ID NO:1+SEQ ID NO:175). SEQ ID NO:1021 had additional mutations E399M, D923K, R1177G on top of the modifications made in SEQ ID NO:636, and SEQ ID NO:783 had the addition of N-terminal EIF4A polypeptide. SEQ ID NO: 1236 had the substitution of STO7D domain instead of CTIP domain, and had significantly improved activity, and SEQ ID NO: 1234 had the further addition of TUDOR peptide and even higher activity in at least one experiment. SEQ ID NO:3028 which applies a different codon optimization strategy also has significantly higher activity than SEQ ID NO:636 and similar or higher activity than SEQ ID NO:1236 in this experiment.

[2155] Notably, for all tested constructs, as assessed by digital PCR, a substantial number of insertion events (~40% for SEQ ID NO:1234 in primary hepatocytes) had full length integrations (ratio of polyA probe to CMV probe).

Example 78: Engineered R2 systems allow integration of different functional transgenes via LNP-mRNA delivery to primary human hepatocytes.

[2156] Engineered R2 systems were tested for their ability to integrate different transgenes (dTomato and Epo) to primary human hepatocytes by LNP-mRNA delivery.

[2157] FIG. 78A is a graph showing results of integration assays using engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOs:636, 618, 783, and 1234, with the R2-PTT dTomato reporter construct SEQ ID NO: 1399. Transfection was carried out with driver and reporter mRNA delivered with LNP to hepatocarcinoma cell line HUH7 and primary human

hepatocyte cells and full-length integration assessed by FACS and dPCR at 72h. The drivers were engineered R2-TG protein constructs described above. The data set included the following controls: Non-treated control, dTomato reporter SEQ ID NO:1399 served as negative control. Error bars indicate standard error of two replicates.

[2158] FIG. 78B is a graph showing results of integration assays using engineered R2-TG drivers having amino acid sequences set forth in SEQ ID NOs:636, 618, 783, 1234. with the R2-PTT erythropoietin (EPO) reporter construct SEQ ID NO: 1400. Transfection was carried out with driver and reporter mRNA delivered with LNP to hepatocarcinoma cell line HUH7 and primary human hepatocyte cells and full-length integration assessed by EPO ELISA and dPCR at 72h. The drivers were engineered R2-TG protein constructs described above. The data set included the following controls: Non-treated control, EPO reporter (SEQ ID NO:1400) negative control. Error bars indicate standard error of two replicates.

[2159] These results demonstrate that engineered R2 systems can integrate full-length functional transgenes in primary hepatocytes using LNP-mRNA delivery.

[2160] SEQ ID NO: 3027 Amino acid sequence of R2 chimeric driver

MASCPKPGPPVSAGAMSLESLTTHSVLAIERGPNLANSGSDFGGGGLGLPLRLLRVS
VGTQTSRSDWVDLVSWSHPGPTSKSQQVDLVSLFPKHRVDLLSKNDQVDLVAQFLPSK
FPPNLAENDLALLVNLEFYRSDLHVECVHKAHWEGLSGLPEVYEQLAPQPCVGETL
HSSLPRDSELFVPEEGSSEKESEDAPKTSPTPGKHGLEQTGEEKVMVTPDKNPPCPCC
GTRVNSVLNLIHLKVSHGKRGVCFRCAKCGKENSNYHSVCHFKKCRGPETEKAPAG
EWICEVCNRDFTTKIGLGQHKRLAHPAVRNQERIVASQPKETSNRGAHKRCWTKEEEE
LLIRLEAQFEGNKNINKLIAEHITTKTAKQISDKRRLLSRKPAEPRMEPGTCHHTTRAA
ASLRTEPEMSHHAQAEDRDNGPGRRPLPGRAAAGGRTMDEIRHPDKANDQQEASSSG
RGPVTRRAAAGKKTKKSIDHPTDKANGQQVAISTRGCLNRGVTAGVETIEGSSHL PDK
DSDGQGTPSREKEGQLQDHYQE VIGAWLS DGRINTFPEAFKGLLGKREIQATINQSILDC
FGGIDSKSQVNKMTRDKKDKKPRPPQKWMKKRAIKRGSYLR FQRLFHLDRGKLAKIV
LDNIECLTCDIPPSEIYSVFKARWETAGKFTSLGDFKFTEKVDNSAFRDLITAKEVEKNM
QEMSKGSAPGPDGITLRDIQKMDPKLSRTTEIFNLWLAAGKIPDGVRECRTVLIPKSTKP
ERLKDINNWRPITIGSTLLRLFSRILTARLSKACPLNPRQKGFQAAGCSENQLLQSIWS
AKREHRPLGVVFDIAKAFD TVSHKHILHVLQRRGVDPHIIGLVSDMYENIHTSITTKKA

RTDRIQIRVGVKQGDPMSPLLFNLAMDPLLCELETNGRGHHRGTNCITAMAFADDLVL
LSDSWEDMKSNIKILEAFCDLTGLKTQGEKCHGFYIKPTKDSYTVNDCAAWTINGTPLN
MILPGESEKYLGLQIDPWTGLAKSDLSTKLDSWLEQIDRVPLKPLQKIELLKTYTIPRLTY
VADHSDLKVSFLETLDQKIRTIVKDWLHLPACTCDAILYSSTKDGGLGITKLAKLIPSVQ
ARRLHRIAQSSDETMMQFLHRDKMEAMYQKLWIQAGGDRENIPSIWEAPPSSEPPNNVS
TNSEWEAPTQKDKFPKPCNWRKNEFKKWTKLASQGRGIVNFERDKISNHWIQYYRRIP
HRKLLTALQLRANVYPTREFLARGRQDQYIKACRHCDADNESCAHIIGNCPVTQDARIK
RHNYICELLLEEAKKKDWVVFKEPHIRDSNKELYKPDILFVKDGRALVVDVTVRYEAA
KSSLEEAAAEKVRKYKHLETEVRHLTNAKDVTFFVGFPLGARGKWHQDNFKLLTELGLS
KSRQVKMAETFTVALFSSVDIVHMFASRARKSMVM

[2161] SEQ ID NO: 13027 DNA sequence of R2 chimeric driver having amino acid
sequence set forth in SEQ ID NO:3027

gaaataagagagaaaagaagagtaagaagaatataagagaattcgccaccatggcttctgcccgaacctggccccgggttcagca
ggggcaatgagttggaaagtgactgaccacccactccgtctcgccatcgaacgtgggcccaattcgttgcaaatccggtacagacttt
gggggggggggtctggggctaccgttacgcctattgagggatcggcgccactcagacctccgctccgactgggtagacctgggtcc
tggagccacccaggacccacgtctaagtcaccagcaggtgacctgggtgtctttatttcctaacacccgggtgacctgtatccaaaaacgac
caggtagacctgggtgctcaattttaccatctaaattcccccaatttggcagaaaatgatttggcttctggtgaacttagagttctacagat
cggatttgcattgtatgagtggttcataaggctgcacattggggagggaatagtggttgcctgagggtgatgaacaacttgcaccacaacc
gtgtgtgggagaaactttacattctagcctccacagacagtgaaactgttgcctgaagaggggagcagcgagaaggagagcgagg
acgcgcaaaaacatctcctccgacgcctgggaacatgttgggaacagactggggaggaaaaagtgatgggtgactgttcctgacaaaa
atccacctgtccttgcgtgtgtaccgggtaaacctgtgtgtaactgtgattgaacatctgaaagtgtcacacgggaaaaaggggggtttgttt
cgggtgtcaaatgtggaaaggaaaatagtaactatcacagtgtgtttgtcattttaagaaatgcaggggtccagagacggagaaagcccc
agctggggagtggttggtaggtatgcaacagagattttacaacaaaattggcctgggacaacacaagagattggcacaccagcagt
gagaaatcaggaaaggatcgttctcccaaccgaaagaacatcaaatagagggtctcacaaaagggtctgacaaaaggaggaggaa
gaattactaataagactggaggctcagttcagggaacaaaaatattaataagcttattgcagaacacataaccacaaaaacagctaagca
gatcagtgacaaaaggcgattgtgtccagaaagccagcagaggagccacgtatggagcctggaacgtgtcatcacaccaggagagca
gctgcgagcctgagaacggagcctgagatgagtcacgcccaggcagaggacagagataatggacctgggagacgcctctgcca
ggcagggcagctgccggaggagagaacaatggacgagataagacgccacctgataaggccaacgatcagcaggaggccagtagcag
tgggagggggccagtgaccaggagagcagctgcaggaaagaaaaaagagattgatcacctactgataaggccaacggcca
gcaggtagccatcagcactagaagggtgtttgaatagaggagtgactgcggggtagaacaatagagggtcagtcacctccctga

taaggacagtgcggacagggcacccccagccgggagaaggagggaattacaagatcactatcaagaagtcacggagcctgggttat
cagatggggaggatcaatacctttccagaggcattcaagggttgtagggaaacgagagatacaggcaacgatcaatcagtcactactgga
ttgtttggggcattgattcaaaaagccaggtgaacaaatgactcgagataagaaagataaaaagccagggagacctcccagaagt
gatgaaaagagagccatcaaaagaggtagctatctcgcttccaacgtctttccatctgtaggggaaaattagccaagattgtttggac
aatatcgaaattgttgacgtcgatataccaccagtgaaattatccgttttaaggccaggtgggaacagcaggcaaatctaccagcttg
gggactttaattactgaaaagggtggacaatagtcattcagggaatttaattacagctaaggaaagtagaaaagaacatgcaagaaatgagc
aaaggctcggctcctggctccagatggaaactctgagagacattcaaaaatggatcctaagctctccaggaccacagagattttcaacctt
tggttgctgcaggcaaaattccggatgggtgagggaaatgcaggactgtcttaattccaaaatcaacaaaccagaacgcttaaggaca
ttaaatttgagaccaattacaatcggatccacctgcttagatttttccaggattttaacagccagactaagcaaggcatgccccctcaac
ccaagacagaagggtcattcaagcagccggatgctcagaaaacctgcaactcctgcaatccataatttggtagcaaaacgggaacac
aggccccctgggtgctgattgttgacatcgcaaaggctttgacacagtaagccacaagcacattctacagttctgcagcggaggggagt
ggatcctcacatcatcgccgtggtgagtgatgtatgaaaacattcatacttctatcaccacgaagaaagcacgtacggatcgatccagat
ccgagttggagttaaacaggcgacccgatgtctcccttattcaatctggcaatggacccccctgctgtgcgaactggaacgaacggc
aggggacaccaccgggcacaaattgcatcacagccatggcgtttgctgacgacctggtcctactgagtgactcctgggaggacatgaag
agcaacatcaagatattggaggccttctgcgatctcaccggcctcaaacgcaggagagaaagtgccacggctttacatcaagccgacg
aaggactctacacgggtcaacgactgtgctgcttgactataaacggcactccccctgaacatgattctccgggtgagtcgagaataacct
gggctgcagatagaccatggagctggttgcaaaatcgacacctctctacaaaactcgactcatggctcgagcaaatcgatagattcca
ctcaaacgctgcagaaaattgaactcctgaaaacatacaccatccctcggtgacatacgtggctgaccattcggacttgaaagttagttcc
tgagacattggatcagaagattcggacaatagttaaagattggtacacctgcctgcatgtacctgcatgccatcctatactctagcacga
aggatggcgccctgggcatcactaaattagcaaagctgattcctagcgtacaggctaggagactacacagaatagcccaatcatcgacg
aaacatgaagcaatttctccaggggacaagatggaaagcaatgtaccagaaactgtggattcaagctggaggggacagagagaacata
ccctcgattgggaagcaccaccgtcgtgagtgaaaccacaaacacgtgagcacaattcggaaatgggaagcaccgaccagaaagataa
atttcaaaagccttgcaattggaggaaaaacgaattcaaaaatggacaaattggcatccaaggccgagggaattgtaaatttgaagag
acaaaattagtaaccattggatccaatactacagacgcatacctcacaggaaactcctcactgactacaactcagggccaacgtttacccc
acgagagaatttctagccaggggttagacaagaccaatacatcaaggcgtgtaggcactgcgatgcggacaacgaatcctgcgcccacat
catcggcaactgccagtgacacaggacgcccgaatcaagggcacaaattacatctgcgaactgcttctcaggaggcgaagaagaag
gactgggtagtgttaagggaaccgcacataagggaattccaacaaggaaactgtacaaacctgacctgatatgtgaaggatggcgctgcac
ttgtcgtggtatgtacagtacggtatgaagcagccaaatcatcgtcggagggaagccgctgcagagaagtgagaagtacaaacacctgg
aaacggaagtaagacatctcacgaatgcaaaggacgttactttgtgggtttccctaggagcgcgggggaaatggccaagataactt
taacttttgactgagcttggcctctccaaatcgaggcaagtgaataatggcagagacttttccacagtagcgctctttcatctgtggacattg
acatatgttgcagtagggccagaaaatctatggttatgtgataataggctgccttctgcggggcttgccttctggccatgcccttctctctc

ctfgcacctgtacctcttggtctttgaataaagcctgagtaggaaggaccggtataaaaaaaaaaaaaaaaaaaaaaaaaaaaaaacgttgact
aagttcttc

[2162] SEQ ID NO: 3028 Amino acid sequence of engineered R2 driver

MAPKKKRVVTVKFKYKGEEKEVDISKIKKVWRVGKMISFTYDDNGKTGRGAVSEKD
APKELLQMLEKSGSETPGTSESATPEGSPKRKVSSAEGAAKEPKRRSARLSAKPPAKVE
AKPKKAAAKDKSSDKKVQTKGKRGAKGKQAEVANQETKEDLPAENGETKTEESPASD
EAGEKEAKSDSGSETPGTSESATPESYEQ LAPQPCVGETLHSSLPRDSELFVPEEGSSEKE
SEDAPKTSPPTPGKHGLEQTGEEKVMVTVPDKNPPCPCGTRVNSVLNIEHLKVSHGK
RGVCFRCAKCGKENSNYHSVVCHFKKCRGPETEKAPAGEWICEVCNRDFTTKIGLGQH
KRLAHPAVRNQERIVASQPKETSNRGAHKRCWTKEEEEELLIRLEAQFEGKNKNINKLIAE
HITTKTAKQISDKRRLLSRKPAEPPREPGTCHHTRRAAASLRTEPEMSHHAQAEDRDN
GPGRRLPGRAAAGGRTMDEIRRHDPDKGNGQQRPTKQKSEEQLQAYYKKTLEERLSAG
ALNTFPRAFKQVMEGRDIKLVINQTAQDCFGCLESISQIRTATRDKKDTV TREKHPKKPF
QKWMKKRAIKKGNLYRFQRLFYLD RGKLAKIILDDIECLSCDIPLSEIYSVFKTRWETT
G
SFKSLGDFKTYGKADNTAFRELITAKEIEKNVQEMSKGSAPGPDGITLGDVVKMDPEFS
RTMEIFNLWLTTGKIPDMVRGCRTVLIPKSSKPDRDKDINNWRPITIGSILLRFLSRIVTAR
LSKACPLNPRQGRFIRAAGCSENKLLQTHIWSAKREHRPLGVVFDIAKAFDTVSHQHII
HALQQREVDPHIVGLVSNMYENISTYITTKRNTHTDKIQIRVGKQGDPMSPLLFNLAM
DPLLCKLEESGKGYHRGQSSITAMAFADDLVLLSDSWENMNTNISILETFCNLTLGLKTQ
GQKCHGFYIKPTKKSYPINDCAAWTINGTPLNMIDPGESEKYLGLQFDPWIGIARSGLST
KLDFWLQRIDQAPLKPLQKTDILKTYTIPRLIYIADHSEVKSALLETLQKIRTAVKEWL
HLPPCTDAILYSSTRDGGLGITKLAGLIPSVQARRLHRIAQSSDDTMKCFMEKEKMEQL
HKKLWIQAGGDRENIPSIWEAPPSSEPPNNVSTNSEWEAPTQKDKFPKPCNWRKNEFKK
WTKLASQGRGIVNFERDKISNHWIQYYRRIPHRKLLTALQLRANVYPTREFLAGRQDQ
YIKACRHCADNESCAHIIGNCPVTQDARIKRHNYICELLLEEAKKDWVVFKEPHIRD
SNKELYKPD LIFVKDGRALVVDVTVRYEAAKSSLEEAAAEKVRKYKHLETEVRHLTNA
KDVT FVG FPLGARGKWHQDNFKLLTELGLSKSRQVKMAETFSTVALFSSVDIVHMFAS
RARKSMVM

[2163] SEQ ID NO:13028 DNA sequence of engineered R2 driver having amino acid sequence set forth in SEQ ID NO:3028

gaaataagagagaaaagaagagtaagaagaatataagagaattcgcaccATGGCCCCTAAGAAAAAGCGGAA
GGTGGTGACCGTGAAGTTCAAGTACAAGGGCGAGGAAAAGGAAGTGGACATCAGC
AAGATCAAGAAGGTCTGGCGGGTGGGCAAGATGATCAGCTTTACCTACGACGATAA
CGGCAAGACAGGCAGAGGCGCCGTGTCCGAGAAGGACGCCCCTAAAGAGCTGCTG
CAGATGCTGGAAAAATCCGGCTCCGAGACTCCTGGAACAAGCGAGAGCGCTACACC
TGAGGGCAGCCCTAAGAGAAAGGTGTCCAGCGCCGAGGGCGCTGCCAAGGAAGAG
CCTAAACGGAGAAGCGCCAGACTGAGCGCCAAGCCCCCGCCAAGGTGGAAGCCA
AGCCTAAGAAGGCCGCCGCTAAGGACAAGAGCAGCGATAAGAAAGTCCAGACCAA
GGGCAAGCGGGGCGCCAAAGGCAACAGGCCGAGGTGGCCAACCAGGAGACAAAG
GAAGATCTGCCTGCTGAGAACGGCGAAACCAAGACCGAGGAATCTCCAGCTTCTGA
CGAGGCCGGAGAGAAGGAGGCCAAAAGCGACAGCGGCTCTGAGACACCTGGAACA
AGCGAGAGCGCCACCCCGGAATCCtacgaacagctggctcccagccctgtgtaggcgagacattacacagcagc
ctgcccagagacagcgagctgttctcgttcggaagagggtctagcgagaaggagagcgaggacgctcctaagcagccctcctacc
ctggcaaacacggcctggagcaaacaggcgaggagaagggtcatgtgaccgtgcccagacaagaacctcctgtctgtgtggcaca
agagtcaacagcgtgtgaacctcatcgagcacctgaaggtgtcccacggcaagagagcggtgtgtcttcagatgcgccaagtgcggcaa
ggagaacagcaactaccacagcgtcgtgtgccaactcaagaagtgcagaggccctgagaccgagaaggctcccgtggcgaatggattt
gcgagggtgtgcaacagagacttcaccaccaagatcggtggccagcacaagcggttagctcacctgtgtgagaaccaggagag
aatcgtggccagccagcccaaggagacaagcaacaggggcgctcacaagagatgtggaaccaaggaggaaggagctgtcatca
gactggagggtcagttcgagggaacaagaacatcaataagcttatcgccgagcacatcaccacaaagaccgccaagcagatcagcgac
aagaggagactgtgtccagaagcccgcgaggagcctagagaggaacctgtgtacatgtcaccacaccagaagagccgcccgtagc
ctgagaaccgagcccgaatgtctcaccacgtcaggccgaggacagagacaacggccctggcaggagacctctccggcagagctg
ctgtgtggaggcagaacctggacgagatcaggagacccccgataagggaacggccagcagagaccacaaagcagaagtccgag
gaacagctccaggcctactacaagaaaacctggaggagagactgtccgagggtctgaacaccttccctagagccttcaacaagt
atggagggcagagacatcaagctgtgtatcaaccagaccgcccaggactgtctggctgctgtgagctatcagccagatcaggaccgc
cacacgggataagaaggacaccgtcaccagagagaagcaccacgaagccctccagaaatggatgaagaaggggccatcaagaa
gggcaattacctgagattcagaggctgttctacctggacagaggcaagctggccaagatcatctggacgacatcgagtgctgagctg
cgacatccctctgagcgagatctacagcgtgttcaagaccagatgggagaccaccggcagcttcaagtctctggcgacttcaagacatac
ggcaaggccgacaacaccgcttcagagagctcatcaccgctaaggagatcgagaagaacgtgcaggagatgagcaagggcagcgct
cctggccccgatggaattacactggcgatgtgtgaagatggaccccgagttcagcagaacaatggagatcttcaacctgtggtcacca
caggcaagatccccgacatggtgagaggctgcagaaccgtgttatcccaagagcagcaagcccacagactgaaggatatcaaca

ctggagaccatcacatcgagcagcatcctgctgcggtgttcagccgattgtgaccgtagactgagcaaacctgtcccctgaacccc
agacagagaggcttcatcagagccgctggctgcagcgagaacctgaagctgtccagacaatcatctggagcgccaagagagagcaca
gaccctggcggtgtttctgtgacatcgccaaggccttcgacaccgtgtccatcagcacattatccacgccctgcagcagaggagggt
cgacctcatattgtgggctgtgtccaacatgtacgagaacatctccacctatattacgaccaagagaaacacccacacagacaagatc
cagatccgggtcggtcaagcaggcgatcctatgtctcctgctgttcaacctcgccatggacctctgctctgcaagctcgaggagtc
tggaagggttaccatagaggccaatcatcgatcaccgcatggccttcgacgacacctggtgttttaagcgacagctggggaacat
gaacaccaacatctctatcctggagacattctgcaaccttacggcctgaaaaacccaggggcagaagtgtcatggcttctacataaccca
ccaagaagagctacacatcaacgactgcgcgcctggacaatcaacggcacccctctcaacatgatcgaccccgcgagctctgagaag
tacctgggctgagttcgatccctggatcgcatcgtagaagcggttgagcaccaagctggacttttggtccaacgcatagaccagg
ccccctgaagcctctcagaaaacggacattttgaaaactatagatcccagactcatctacatcgccgaccactccgaggtcaagagc
gctctgctggaacccctggaccagaaaatcagaacccgctcaaggaaatggctgcacctgccccctgcacatcgacgctatcctgtac
agcagcaccaggcgacggcgttaggaatcaccaagttagcggcctcatcccagcgtgcaggctaggagactccacagaatcgccc
agagcagcgacgacaccatgaagtgtctcatgagagaaggaaaagatggagcagctgcacaagaagctgtggatccaacggcgaggag
atcgtgagaacattcccagcatctggaggccctcccagctcagaacctcccataacgtgtccaccaattctgaatggagggtctccac
ccagaagataagtttccaagccctgcaactggcggaagaacgagttcaaaaatggacaagctggctagccagggcaggggcatc
gtgaacttcgaaaggacaagattagcaaccactggatccagtactacagaagaatccccacagaagctcctcacggccctccagttac
gtgcaaacgtgtaccacaaccagagaattctggccagaggcagacaggaccagtacatcaaggcctgccggcactgcgatccgataac
gagagctgtgtcacatcatcggaactgccccgtcaccaggacgccaagaatcaagagacacaactacatctcgagttacttctggag
gaagccaagaagaagattgggtgtgttcaaggagcccccacatcagagactccaacaaggagctgtacaagccagacctcatcttctg
aaggacggcagagccctgggtgggacgtcacctgagatagaggctgctaagtccagccttggaggagctgcgcgagagaagtgc
ggaagtacaagcacctggaaacagaggtcagacacctaccaacgccaaggacgttaccttctgtggccttccccctaggtgtagaggc
aaatggcaccaggacaactcaagctgcttaccgagctgggctctccaagagccgacaggttaagatggccgagaccttcagaccgtc
gcctgttcttctgttgacattgtgcacatgttcgctccagagccagaagagcatgtgtatgtgataataggctgccttctgctgggctt
gccttctggccatgcccttcttctcccttgacacctgtaccttctgttgaataaagcctgagtaggaaggaccggtaaaaaaaaaa
aaaaaaaaaaaaaaaaacgttgactaaa
aaaaaaaaaaaaa

[2164] SEQ ID NO:3029 Amino acid sequence of engineered R2 driver

MAPKKKRVKSASQDSRSRDNGPDGMEPEGVIESNWNIVDSFDDMNLSESLLRGIYAY
GFEKPSAIQQRAILPCIKGYDVIAQAQSGTGKTATFAISILQQIELDLKATQALVLAPTRE
LAQQIQKVVMALGDYMGASCHACIGGTNVRAEVQKLQMEAPHIIVGTPGRVFDMLNR
RYLSPKYIKMFVLDEADEMLSRGFKDQIYDIFQKLNSNTQVVLLSATMPSDVLEVTCKF

MRDPIRILVKEELTLEGIRQFYINVEREEWKLDLTCDLYETLTITQAVIFINTRRKVDWL
TEKMHARDFTVSAMHGDMQKQKRDVIMREFRSGSSRVLITTDLLARGIDVQQVSLVIN
YDLPTNRENYIHRIGRGGFRGRKGVAINMVTEEDKRTLRIETFYNTSIEEMPLNVADLI
SGSETPGTSESATPEGSVTVKFKYKGEEKEVDISKIKKVVWRVGMISFTYDDNGKTGRG
AVSEKDAPKELLQMLEKSGSETPGTSESATPEGSPKRKVSSAEGAAKEPKRRSARLSA
KPPAKVEAKPKKAAAKDKSSDKKVQTKGKRGAKGKQAEVANQETKEDLPAENGETK
TEESPASDEAGEKEAKSDSGSETPGTSESATPESYEQ LAPQPCVGETLHSSLPRDSELFVP
EEGSSEKESDAPKTSPPTPGKHGLEQTGEEKVMVTVDPKNPPCPCGTRVNSVLNIEH
LKVSHGKRGVCFRCAKCGKENSNYHSVCHFKKCRGPETEKAPAGEWICEVCNRDFTT
KIGLGQHKRLAHPVRNQRIVASQPKETSNRGAHKRCWTKEEEELLIRLEAQFEGNKN
INKLIAEHITTKTAKQISDKRRLSRKPAEEPRMEPGTCHHTRRAAASLRTEPEMSHHAQ
AEDRDNGPGRRPLPGRAAAGGRTMDEIRRHDPKGNGQQRPTKQKSEEQLQAYYKKT
EERLSAGALNTFPRAFKQVMEGRDIKLVINQTAQDCFGCLESISQIRTATRDKKDVTRE
KHPKKPFQKWMKKRAIKKGNLYLRFQRLFYLD RGKLAKIILDDIECLSCDIPLSEIYSVFK
TRWETTGSFKSLGDFKTYGKADNTAFRELITAKEIEKNVQEMSKGSAPGPDGITLGDVV
KMDPEFSRTMEIFNLWLTGKIPDMVRGCRTVLIPKSSKPDRLKDINNWRPITIGSILLRL
FSRIVTARLSKACPLNPRQGRGFIRAAGCSENKLLQTIWSAKREHRPLGVVFVDIAKAF
DTVSHQHIIHALQQREVDPHIVGLVSNMYENISTYITTKRNTHTDKIQIRVGVKQGDPM
PLLFNLAMDPDLLCKLEESGKGYHRGQSSITAMAFADDLVLLSDSWENMNTNISILETFC
NLTLGLKTQGGKCHGFYIKPTKKS YTINDCAAWTINGTPLNMIDPGESEKYLGLQFDPWI
GIARSGLSTKLDFWLQRIDQAPLKPLQKLDILKTYTIPRLIYIADHSEVKSALETL DQKIR
TAVKEWLHLPCTCDAILYSTRDGG LGITKLAGLIPSVQARRLHRIAQSSDDTMKCFM
EKEKMEQLHKKLWQAGGDRENIPSIWEAPPSSEPPNNVSTNSEWEAPTQKDKFPKPCN
WRKNEFKKWTKLASQGRGIVNFERDKISNHWIQYYRRIPHRKLLTALQLRANVYPTRE
FLARGRQDQYIKACRHCDADNESCAHIIGNCPVTQDARIKRHNYICELLLEEAKKKDWV
VFKEPHIRDSNKELYKPD LIFVKDGRALVVDVTVRYEAAKSSLEEAAA EKVRKYKHLE
TEVRHLTNAKDVT FVGFP LGARGKWHQDNFKLLTELGLSKSRQVKMAETFSTVALFSS
VDIVHMFASRARKSMVM

[2165] SEQ ID NO:13029 DNA sequence of engineered R2 driver having amino acid
sequence set forth in SEQ ID NO:3029

gaaataagagagaaaagaagagtaagaagaaatataagagaattcgccaccATGGCCCCTAAGAAAAAGCGGAA
GGTGAGCGCCAGCCAGGACAGCAGAAGCCGGGACAACGGCCCTGACGGAATGGAG
CCTGAGGGAGTGATCGAATCCAACCTGGAACGAGATCGTGGACAGCTTCGACGACAT
GAACCTGAGCGAGAGCCTGCTGAGAGGCATCTACGCTTACGGCTTCGAGAAGCCTA
GCGCCATCCAGCAAAGAGCAATCCTGCCATGCATCAAGGGCTACGACGTCATCGCC
CAGGCCCAGAGCGGCACCGGCAAAACCGCCACCTTCGCCATCTCTATCCTGCAGCA
GATCGAGCTGGATCTGAAAGCCACACAAGCGTTGGTGCTGGCCCCCACAAGAGAGC
TGGCCCAGCAGATCCAGAAGGTGGTGATGGCTCTGGGCGACTACATGGGCGCCTCT
TGTCATGCCTGCATCGGCGGCACCAATGTGCGGGCTGAAGTGCAGAAGCTCCAGAT
GGAAGCTCCCCACATCATTTGTGGGAACACCAGGCAGGGTGTTTCGATATGCTGAATC
GCCGGTACCTGTCTCCTAAGTATATCAAGATGTTTCGTGCTGGACGAGGCCGACGAA
ATGCTGTCTAGAGGTTTCAAGGACCAGATCTACGACATTTTTCAGAACTGAACAGC
AACACCCAGGTTGTGCTGCTGTGTCAGCCACCATGCCTAGCGATGTGCTCGAGGTGACC
AAGAAGTTCATGCGGGACCCTATCAGAATCCTGGTGAAGAAAGAGGAGCTGACTCT
GGAGGGCATCAGACAGTTCTACATCAACGTGGAAAGAGAGGAATGGAAGCTGGAC
ACCCTGTGCGACCTGTATGAGACACTGACCATCACCCAGGCCGTGATCTTTATCAAT
ACCAGAAGAAAGGTCTGACTGGCTGACAGAGAAGATGCACGCCCCGGGACTTCACCGT
GTCCGCCATGCACGGCGATATGGACCAAAAGGAACGGGATGTGATTATGAGAGAGT
TCCGGTCCGGCAGCAGCCGGGTGCTGATCACAACAGATCTGCTGGCTAGAGGAATC
GATGTGCAGCAGGTGAGCCTGGTCATCAACTACGATCTGCCTACCAACAGAGAAAA
TTACATCCACAGAATCGGCAGAGGCGGAAGATTTCGGCAGAAAGGGCGTGGCCATA
AACATGGTGACAGAGGAAGATAAGCGGACCCTTAGAGACATCGAAACCTTTTACAA
CACCAGCATCGAGGAAATGCCCCTGAACGTGGCCGACCTGATCTCCGGCTCCGAGA
CTCCTGGAACAAGCGAGAGCGCTACACCTGAGGGCAGCGTGACCGTGAAGTTCAAG
TACAAGGGCGAGGAAAAGGAAGTGGACATCAGCAAGATCAAGAAGGTCTGGCGGG
TGGGCAAGATGATCAGCTTTACCTACGACGATAACGGCAAGACAGGCAGAGGCGCC
GTGTCCGAGAAGGACGCCCCCTAAAGAGCTGCTGCAGATGCTGGAAAAATCCGGCTC
CGAGACTCCTGGAACAAGCGAGAGCGCTACACCTGAGGGCAGCCCTAAGAGAAAG
GTGTCCAGCGCCGAGGGCGCTGCCAAGGAAGAGCCTAAACGGAGAAGCGCCAGAC
TGAGCGCCAAGCCCCCGCCAAGGTGGAAGCCAAGCCTAAGAAGGCCGCCGCTAA
GGACAAGAGCAGCGATAAGAAAGTCCAGACCAAGGGCAAGCGGGGCGCCAAAGGC
AAACAGGCCGAGGTGGCCAACCAGGAGACAAAGGAAGATCTGCCTGCTGAGAACG

GCGAAACCAAGACCGAGGAATCTCCAGCTTCTGACGAGGCCGAGAGAAGGAGGC
CAAAAGCGACAGCGGCTCTGAGACACCTGGAACAAGCGAGAGCGCCACCCCGAA
TCCtacgaacacgtggtccccagccctgtgttaggcgagacattacacagcagcctgcccagagacagcgagctgttcgtccggaag
agggtcttagcagagaaggagagcgaggacgtcctaagaccagccctcctacccctggcaaacacggcctggagcaaacaggcgagg
agaaggtcatgtgaccgtgcccagacaagaacctccctgtccttgctgtggcacaagagtcaacagcgtgctgaacctcatcgagcacct
gaaggtgtccacggcaagagaggcgtgtcttcagatgcgccaagtgcggcaaggagaacagcaactaccacagcgtcgtgtgcccac
ttcaagaagtgcagaggccctgagaccgagaaggctcccgctggcgcaatggattgaggtgtgcaacagagacttcaccaccaagat
cggcctggggccagcacaagcgttagctcacctgctgtgagaacacaggagagaatcgtggccagccagcccaaggagacaagcaa
cagggcgctcacaagagatgctggaccaaggaggaaagaggagctgctcatcagactggaggctcagttcaggggcaacaagaacat
caataagcttatcgccgagcacatcaccacaaagaccgccaagcagatcagcgacaaggagagactgctgtccagaagcccgcgag
gagcctagaatggaacctgtgtacatgtcaccacaccagaagagccgcccgtagcctgagaaccgagcccgaatgtctcaccacgtcta
ggccgaggacagagacaacggccctggcaggagacctctcccgagagctgctgtgagggcagaacctggacagatcagga
gacaccccgataagggaacggccagcagagaccacaaagcagaagtccgaggaacagctccaggcctactacaagaaaacctgg
aggagagactgtccgaggagctctgaacaccttccttagagccttcaacaagtcatggagggcagagacatcaagctggtgatcaacc
agaccgcccaggactgcttcggctgcttgagcttatcagccagatcaggaccgccacacgggataagaaggacaccgtcaccagaga
gaagcaccccaagaagccctccagaatggatgaagaaggggccatcaagaagggaattacctgagattccagaggctgtttctacct
ggacagaggcaagctggccaagatcatctgacgacatcagtgctgagctgcgacatccctctgagcgagatctacagcgtgttcaa
gaccagatgggagaccaccggcagcttcaagtctctggcgacttcaagacatacggcaaggccgacaacaccgcttcagagagctca
tcaccgctaaggagatcgagaagaacgtgcaggagatgagcaagggcagcgtcctgccccgatggaattactggcgatgtggt
gaagatggaccccgagttcagcagaacaatggagatcttcaacctgtggctcaccacaggcaagatccccgacatggtgagagggtgca
gaacctgtcttatccccaagcagcagaagcccgacagactgaaggatatacaactggagacccatcaccatcggcagcatcctgctg
cggctgttcagccgattgtgaccgctagactgagcaaacctgtccctgaacccagacagagaggcttcacagagccgctggtgctg
agcgagaacctgaagctgctcagacaatcatctggagcgccaagagagagcacagacccctggcggtgtttctgtgacatcgccaa
ggccttcgacaccgtgtcccatcagcacattatccacgccctgcagcagaggagggtcgacctcatattgtgggcctggttccaacatgt
acgagaacatctccacctatattacgaccaagagaaacacccacacagacaagatccagatccgggtcggcgtaagcaggggcgatcct
atgtctcctctgtgttcaacctgccaatggacctctgctctgcaagctcgaggagtctggaagggtctaccatagaggccaatcatcgatc
accgcatggccttcgacgacgtggtgttfaagcgacagctgggagaacatgaaccaacatctctatcctggagacattctgcaa
ccttaccggcctgaaaaccaggggccagaagtgtcatggttctacatcaaacccaccaagaagagctacaccatcaacgactgcgccgc
ctggacaatcaacggcaccctctcaacatgatgaccccgcgagtctgagaagtacctgggcctgcagttcgatccctggatcggcatc
gtagaagcggtgtagcacaagctggacttttgctccaacgcatagaccaggccccctgaagcctcttcagaaactggacattttga
aaacttatacgatccccagactcatctacatcgccgaccactccgaggtcaagagcgctctgctggaacccctggaccagaaaatcagaac
cgccgtcaaggaaatggctgcacctgcccccttgacatgcgacgtatcctgtacagcagcaccaggggacggcggttaggaatcacca

agttagccggcctcatcccagcgtgcaggctaggagactccacagaatgccagagcagcgacacaccatgaagtgttcattggag
aaggaaaagatggagcagctgcacaagaagctgtggatccaagcgggcgagatcgtgagaacattccagcatctgggaggccctc
ccagctcagaacctccaataacgtgtccaccaattctgaatgggaggctcccaccagaaggataagttccaagccctgcaactggcg
gaagaacgagttcaaaaaatggacaaagctggctagccagggcaggggcatcgtgaactcgaagggaagattagcaaccactgg
atccagtactacagaagaatccccacagaagctcctcaccgccctccagttacgtgcaaacgtgtaccaaccagagaattcctggcca
gaggcagacaggaccagtacatcaaggcctgccggcactgcgatgccgataacgagagctgtgctcacatcatcggaactgccccgtc
accagcagcggccagaatcaagagacacaactacatctgcgagttacttctggaggaagccaagaagaaggattgggtgtgtcaaggag
ccccacatcagagactccaacaaggagctgtacaagccagacctcatcttctgaaggacggcagagccctgggtgggacgtcacgt
cagatacagggctgctaagtccagccttgaggaggctgccgccgagaaagtgcggaagtacaagcacctggaacagaggtcagacac
ctcaccaacccaaggacgttaccttctgtgggcttccccctaggtgctagaggcaaatggcaccaggacaactcaagctgcttaccgagc
tgggcctctccaagagccgacaggttaagatggccgagaccttcagcacctgcgccctgttcttctgtggacattgtgcacatgttcgcct
ccagagccagaaagagcatggtgatgtgataataggctgccttctgcgggcttgcccttctggccatgcccttcttctccttgccctgta
cctcttggtcttgaataaagcctgagtaggaaggaccggtaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaacgttgactaaaaaaaaa
aaa

CLAIMS

1. A nucleic acid encoding an engineered driver comprising an R2 protein or portion thereof fused to at least one heterologous polypeptide, wherein:
 - a. the at least one heterologous polypeptide comprises a heterologous nucleosome binding polypeptide, a heterologous nucleic acid binding polypeptide, a heterologous RNA/DNA repair polypeptide, a heterologous RNA/DNA processing polypeptide, a heterologous localization signal, a heterologous reverse transcriptase, a heterologous endonuclease, a heterologous intrinsically disordered polypeptide, or any combination thereof; and
 - b. the engineered driver exhibits an improvement in at least efficiency, compared to a reference R2 protein.
2. The nucleic acid of claim 1, wherein the at least one heterologous polypeptide is fused to the N-terminus of the R2 protein or portion thereof.
3. The nucleic acid of claim 1, wherein the at least one heterologous polypeptide is fused to the C-terminus of the R2 protein or portion.
4. The nucleic acid of claim 1, wherein the at least one heterologous polypeptide is fused internally within the R2 protein or portion thereof, or a combination thereof.
5. The nucleic acid of any one of claims 1-4, wherein at the least one heterologous polypeptide is a heterologous nucleosome binding polypeptide.
6. The nucleic acid of claim 5, wherein the heterologous nucleosome binding polypeptide is a HMGN1 polypeptide, a HMGB1 polypeptide, or a StkC DNA binding domain.
7. The nucleic acid of claim 6, wherein the heterologous nucleosome binding polypeptide is a HMGN1 polypeptide.
8. The nucleic acid of claim 7, wherein the HMGN1 polypeptide comprises an amino acid sequence set forth in SEQ ID NO: 40, or an amino acid sequence that is at least 70% identical thereto.
9. The nucleic acid of claim 8, wherein the HMGN1 polypeptide comprises at least one amino acid substitution, wherein the at least one amino acid substitution is: D100E or D100G, relative to SEQ ID NO: 40.
10. The nucleic acid of any one of claims 1-9, wherein the at least one heterologous polypeptide comprises a heterologous RNA/DNA repair polypeptide.

11. The nucleic acid of claim 10, wherein the at least one heterologous RNA/DNA repair polypeptide is a CTIP polypeptide.
12. The nucleic acid of claim 10, wherein the at least one heterologous RNA/DNA repair polypeptide comprises a CTIP polypeptide.
13. The nucleic acid of claim 11 or claim 12, wherein the CTIP polypeptide comprises an amino acid sequence set forth in SEQ ID NO: 61 or an amino acid sequence that is at least 70% identical thereto.
14. The nucleic acid of claim 13, wherein the CTIP polypeptide comprises at least one amino acid substitution, wherein the at least one amino acid substitution is: T15I, R123K, P181L, M202L, M202F, M202T, A237L, A237I, A237N, A237K, N288D, D18F, I49C, S17D, D18C, Q46T, I2A, C198S, E155S, L281G, I186S, Q46K, or a combination thereof relative to SEQ ID NO: 61.
15. The nucleic acid of claim 10, wherein the at least one heterologous RNA/DNA repair polypeptide comprises a heterologous PIP motif polypeptide, and the heterologous PIP motif polypeptide is p21, FEN1 or CHAF1A.
16. The nucleic acid of claim 15, wherein, the R2 protein or portion thereof lacks a functional native PIP motif.
17. The nucleic acid of any one of claims 1-16, wherein the at least one heterologous polypeptide comprises a heterologous RNA/DNA processing polypeptide.
18. The nucleic acid of claim 17, wherein the heterologous RNA/DNA processing polypeptide is a RNA helicase polypeptide, a RNA helicase recruitment motif polypeptide, or a RAD51 polypeptide.
19. The nucleic acid of claim 18, wherein the heterologous RNA/DNA processing polypeptide is a RNA helicase polypeptide.
20. The nucleic acid of claim 19, wherein the RNA helicase polypeptide is a EIF4A polypeptide.
21. The nucleic acid of claim 20, wherein the EIF4A polypeptide comprises an amino acid sequence set forth in SEQ ID NO: 585, or an amino acid sequence that is at least 70% identical thereto.
22. The nucleic acid of claim 19, wherein the RNA helicase polypeptide is a Dengue virus RNA helicase polypeptide.

23. The nucleic acid of claim 22, wherein the Dengue virus RNA helicase polypeptide comprises an amino acid sequence set forth in SEQ ID NO: 1875, or an amino acid sequence that is at least 70% identical thereto.
24. The nucleic acid of claim 18, wherein the at least one heterologous polypeptide comprises a RNA helicase recruitment motif polypeptide.
25. The nucleic acid of claim 24, wherein the RNA helicase recruitment motif polypeptide is a TUDOR domain polypeptide.
26. The nucleic acid of claim 25, wherein the TUDOR domain polypeptide comprises an amino acid sequence set forth in SEQ ID NO: 584, or an amino acid sequence that is at least 70% identical thereto.
27. The nucleic acid of claim 26, wherein the TUDOR domain polypeptide comprises an amino acid sequence set forth in SEQ ID NO: 1853, or an amino acid sequence that is at least 70% identical thereto.
28. The nucleic acid of claim 24, wherein the RNA helicase recruitment motif polypeptide is a canine hepacivirus core protein polypeptide.
29. The nucleic acid of claim 28, wherein the canine hepacivirus core protein polypeptide comprises an amino acid sequence set forth in SEQ ID NO: 789, or an amino acid sequence that is at least 70% identical thereto.
30. The nucleic acid of claim 18, wherein the at least one heterologous polypeptide comprises a RAD51 polypeptide.
31. The nucleic acid of claim 30, wherein the RAD51 polypeptide comprises an amino acid sequence set forth in SEQ ID NO: 21, or an amino acid sequence that is at least 70% identical thereto.
32. The nucleic acid of any one of claims 1-31, wherein the at least one heterologous polypeptide comprises a heterologous nucleic acid binding polypeptide.
33. The nucleic acid of claim 32, wherein the heterologous nucleic acid binding polypeptide is a non-sequence specific DNA binding polypeptide.
34. The nucleic acid of claim 33, wherein the non-sequence specific DNA binding polypeptide comprises a Sto7d DNA binding domain polypeptide, a Sso7d DNA binding domain polypeptide, or a UL42 DNA binding domain polypeptide.

35. The nucleic acid of claim 34, wherein the Sto7d DNA binding domain polypeptide comprises an amino acid sequence set forth in SEQ ID NO: 42, or an amino acid sequence that is at least 70% identical thereto.
36. The nucleic acid of claim 34, wherein the Sso7d DNA binding domain polypeptide comprises an amino acid sequence set forth in SEQ ID NO: 23, or an amino acid sequence that is at least 70% identical thereto.
37. The nucleic acid of claim 34, wherein the UL42 DNA binding domain polypeptide comprises an amino acid sequence set forth in SEQ ID NO: 587 or an amino acid sequence that is at least 70% identical thereto.
38. The nucleic acid of any one of claims 1-37, wherein the at least one heterologous polypeptide is a heterologous intrinsically disordered region polypeptide.
39. The nucleic acid of claim 38, wherein the heterologous intrinsically disordered region polypeptide comprises an amino acid sequence set forth in any one of SEQ ID NO: 1171-1216 or an amino acid sequence that is at least 70% identical thereto.
40. The nucleic acid of claim 38 or claim 39, wherein the R2 protein or portion thereof has a deletion in a native intrinsically disordered region.
41. The nucleic acid of any one of claims 1-40, wherein the at least one heterologous polypeptide comprises a heterologous localization signal.
42. The nucleic acid of claim 41, wherein the heterologous localization signal is a nuclear localization signal (NLS).
43. The nucleic acid of claim 42, wherein the NLS comprises an amino acid sequence set forth in any one of SEQ ID NOs: 110-112, 116, or 1854.
44. The nucleic acid of claim 41, wherein the localization signal is a nucleolar localization signal (NoLS).
45. The nucleic acid of claim 44, wherein the NoLS comprises an amino acid sequence set forth in any one of SEQ ID NOs: 113-115.
46. The nucleic acid of any one of claims 1-45, wherein the R2 protein or portion thereof comprises a reverse transcriptase domain and an endonuclease domain.
47. The nucleic acid of any one of claims 1-46, wherein the R2 protein is a wild type (WT) R2 protein.

48. The nucleic acid of any one of claims 1-46, wherein the R2 protein or portion thereof comprises at least one amino acid substitution, at least one amino acid deletion, at least one amino acid addition, or a combination thereof, relative to a WT R2 protein.
49. The nucleic acid of claim 48, wherein the R2 protein or portion thereof has at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% sequence identity with the WT R2 protein.
50. The nucleic acid of any one of claims 47-49, wherein the WT of the R2 protein comprises an amino acid sequence set forth in any one of SEQ ID NOS: 1-2, 401-480, and 30001-30060.
51. The nucleic acid of any one of claims 48-50, wherein the R2 protein or portion thereof comprises at least one amino acid substitution, relative to the WT of the R2 protein.
52. The nucleic acid of claim 51, wherein the at least one amino acid substitution stabilizes the reverse transcriptase domain of the R2 protein and/or the association of the reverse transcriptase domain with an RNA and/or a DNA, relative to the reverse transcriptase domain of the WT of the R2 protein.
53. The nucleic acid of claim 51 or claim 52, wherein the at least one amino acid substitution is relative to SEQ ID NO: 1.
54. The nucleic acid of claim 53, wherein the at least one amino acid substitution is a substitution in at least one of positions F148, K700, F667, C145, P280, D555, T1011, I1219, and A1282 relative to SEQ ID NO:1.
55. The nucleic acid of claim 53 or claim 54, wherein the at least one amino acid substitution position P280 relative to SEQ ID NO:1.
56. The nucleic acid of claim 53 or claim 54, wherein the at least one amino acid substitution is a substitution in position D555 relative to SEQ ID NO:1.
57. The nucleic acid of claim 53 or claim 54, wherein the at least one amino acid substitution is a substitution in position T1011 relative to SEQ ID NO:1.
58. The nucleic acid of claim 53 or claim 54, wherein the at least one amino acid substitution is a substitution in position I1219 relative to SEQ ID NO:1.
59. The nucleic acid of claim 53 or claim 54, wherein the at least one amino acid substitution is a substitution in position A1282 relative to SEQ ID NO:1.
60. The nucleic acid of claim 53, wherein the at least one amino acid substitution is a substitution selected from the group consisting of L18P, E19L, S26N, V27R, A29T, I30A,

R32G, N35I, S36R, L37K, A38R, L49F, G209R, K289R, L316H, E399M, E472G, Y477H, T514R, D537Y, T538K, H544Q, D555K, N562D, W762R, R822K, S895K, Q909N, R964K, G966N, T989Y, I1003M, T1011S, A1012G, L1013R, T1016L, Q1019R, E1027D, P1033R, C1079K, K1085R, L1089R, H1090Y, I1219N, A1282G, T989I, K579R, K547R, V660I, T608P, F956I, I1003L, T1016S, T680S, L594P, P52R, P418S, P78G, L51G, P78V, P180S, V539G, P180K, P437R, V539R, N640A, W677C, V687W, R866I, A747G, W762E, T908I, W959K, D1037N, Y1194L, R1292P, K1210R, K993R, K755R, C518R, K314R, K1182R, K385M, P437M, T448Y, C518Y, E1008R, T305R, P280K, A1012R, N269R, L213K, E214K, Q215K, E214R, T1011R, E1015R, L213R, G257M, G209F, W762D, Y1194L, V687W, W677C, R1292P, T908I, W959K, N640A, Q909N, R866I, A1282G, W762E, W959K, T908I, Y579R, T1011A, Y571H, E630D, D1075E, L1089M, A442P, H423Q, E173D, P188L, C145G, N301G, R355K, G257R, V251G, T603A, N363H, P1033S, F667Y, F1365L, V501L, G257R+C145G, W762R+C145G, G257R+E173D, N363H+E173D, A442P+E173D, W762R+E173D, G257R+P188L, W762R+P188L, G257R+V251G, S895K+V251G, N363H+G257R, R568S, E542K, P985R, G561R, Y563R, L1184K, H1030R, A557R, I558R, Q567R, N271W, E350R, L316R, F304K, D923K, A317W, E1010R, Q1021R, E419R, T1011F, E769R, W345R, V110R, R568G, F135G, N104Y, V142Y, P115K, A111Y, L129R, V98R, D126R, E144R, Q112R, Y136R, F148K, T1016K, K700D, E188D, E188A, R435K, L438A, A1191L, P1195G, P1195G, R1197K, L1200I, G1331A, G1335A, P593G, S595V, E596S, S599V, P593A, S595A, E596A, S599A, K612D, D616K, K612A, D616A, K856D, D6860K, K856A, D6860A, N891S, T892V, S895V, N891A, T892A, S895A, E888S, E898S, T899V, N902S, E888A, E898A, T899A, N902A, T934V, G937P, P939G, T934A, G937A, K1252D, D1255K, W1256G, K1252A, D1255A, W1256A, A877V, A877I, A877M, F876Y, A877H, Q955R, L952I, T989L, E947S, G658R, Q642R, F973R, G647T, D990V, D990R, N941K, V727L, M874L, D227W, D227F, C344V, R669V, V600T, R866L, T989V, M449E, T679L, Y478L, M671S, F956V, M686S, W1126P, T1367R, C404S, E1363K, F973Q, M420T, R669F, G689E, N1219L, K480R, K508Q, E812K, V1368R, R1157M, G217P, H1320E, N891K, K1269G, K1269G, D1255G, T670L, T670L, H1180Q, K907R, T996A, I778F, R574G, K1151H, E194G, R574S, D1288A, F186G, T969E, G432Q, D591Q, M553R, T174K, W1028K, D534Y, D591W, D972R, W1028Y, D972Q, P169R, L975R, C587K, I978K, D591R, D182R, D972M,

A1024K, C170K, P985K, G172R, S178R, G172K, S177R, A166R, S177W, P548R, S204R, H544R, G575H, F1135T, N1123K, P1116D, P998H, S645Q, P998Q, D879A, D879M, K576M, D879W, E643R, T994W, N1140R, D1133R, D583R, M644R, G575R, A1128R, A649R, E1125R, F1135Y, C750K, P1129R, F1135R, L1057R, H1006K, L582K, D573R, T994R, D878R, A578K, D839R, G311S, T532K, T728K, P1116K, L243K, E247K, L249F, K255R, V258I, F260Y, Y272H, E286K, P291A, D303S, I308R, I327L, K333R, G339R, I370L, I374L, K377R, S383K, L474I, T481I, L482I, F498L, M502L, Q516R, F519I, G520K, C521L, S526K, D534K, L572K, S589R, F611Y, P545K, F549R, L564R, L569R, F570Y, D573N, K377R, L582I, I585K, I597V, F601Y, F628L, L631P, V641L, A649S, P650A, V661K, E666K, E672R, F674Y, W677I, V687L, C690S, N738H, Q741R, L754I, A764S, H768R, L771I, V776I, F846Y, L848I, I896L, W933F, Y995F, A1038G, A1063L, I778F, V785I, L794M, E798G, K821G, K821R, V835I, L844I, L854I, C855R, H865K, C901L, D929N, A932P, M942L, K950R, Y951H, D957N, S965P, T969E, P998S, Y1002H, L1014I, T1023K, E1027K, Y1041H, S1043R, K1053R, V1061R, Q1062R, T1130K, Q1173R, Q1233K, F1278L, V1287I, V1291I, E1317Q, V1318I, M1361L, E1363K, R1064K, K1078R, F1080L, E1082K, K1092R, G1098R, E1127K, E1127Q, K1132R, Q1155H, F1162Y, W1171F, L1184I, Q1188K, C1215Y, N1228Q, E1249K, V1257T, H1313P, F1345N, K1355R, K1386R, V727M, T759G, Q795R, I814C, T819E, H825K, Q868G, Q868D, A931P, D944G, F973S, L175S, A1012S, I683L, S699G, I707P, W710Y, T714A, I718T, I726V, N813G, V881T, L906M, S648K, S699K, S717P, A748R, D800P, V239M, E334A, T532G, K622Q, T1367R, H1378R, F1380Y, Y817R, Q911K, Q911S, L657Y, L720Y, A779S, K780N, D199K, A779K, D783G, D851N, E949Y, E949Y, V299E, E949Y, G520A, D591G, S599D, T1122L, L1285I, N1103D, P739R, E638L, N889G, Y1242K, E400M, D1133P, S645K, V639T, I745R, G772A, A747S, A748P, I778F, M841L, M850L, E949L, F983E, R669L, L213K, L564Q, F973E, R669I, F667G, and F973E, or combinations thereof relative to SEQ ID NO: 1.

61. The nucleic acid of claim 53, wherein the at least one amino acid substitution is a substitution selected from the group consisting of E399M, Y477H, D555K, W762R, R822K, H1090Y, F956I, P78V, W677C, K314R, K385M, P437M, T448Y, C518Y, T305R, P280K, N269R, L213K, E214K, Q215K, T1011R, R435K, L438A, R1197K, G1331A, T989L, Q955R, L952I, E947S, Q642R, G647T, N941K, V727L, M874L,

D227W, D227F, C344V, R669V, V600T, T989V, M449E, T679L, Y478L, M671S, F956V, M686S, W1126P, T1367R, C404S, E1363K, K508Q, E812K, V1368R, R1157M, N1123K, P1116D, P998H, S645Q, P998Q, N1123K, P1116D, P998H, S645Q, P998Q, D879A, D879M, K576M, G311S, T532K, L572K, S589R, F611Y, F846Y, L848I, I896L, W933F, Y995F, A1038G, A1063L, T1130K, Q1173R, Q1233K, F1278L, V1287I, V1291I, E1317Q, V1318I, M1361L, E1363K, K1386R, V727M, T759G, Q795R, I814C, T819E, H825K, Q868G, Q868D, A931P, D944G, F973S, L175S, A1012S, I683L, S699G, I707P, W710Y, T714A, I718T, I726V, N813G, V881T, L906M, S648K, S699K, S717P, A748R, D800P, V239M, E334A, T532G, K622Q, D923K, P739R, E638L, N889G, Y1242K, E400M, T1011S, T1011A, I1219N, A1282G, L213R, A1012G, E542K, G257R, C145G, L316R, D537Y, D1133P, and K1085R, or combinations thereof relative to SEQ ID NO: 1.

62. The nucleic acid of claim 53, wherein the at least one amino acid substitution comprises the following substitutions: P280K, D555K, T1011S, I1219N, and A1282G relative to SEQ ID NO:1.
63. The nucleic acid of claim 53, wherein the at least one amino acid substitution is a substitution in at least one of positions G257R, P280K, N363H, E399M, V539G, H423Q, A442P, D537Y, K547R, D555K, R669L, N889G, F973E, T1011A, Q1019R, D1133P, K1085R, R1177G, Y1242K, A1282G relative to SEQ ID NO:1.
64. The nucleic acid of claim 63, wherein the at least one amino acid substitution comprises the following substitutions: G257R, P280K, N363H, E399M, V539G, H423Q, A442P, D537Y, K547R, D555K, R669L, N889G, F973E, T1011A, Q1019R, D1133P, K1085R, R1177G, Y1242K, and A1282G relative to SEQ ID NO:1
65. The nucleic acid of claim 53, wherein the engineered driver comprises an amino acid sequence set forth in any one of SEQ ID NOs: 230-238, 668, and 1378, or an amino acid sequence that is at least 70% identical thereto.
66. The nucleic acid of claim 51 or claim 52, wherein the at least one amino acid substitution is relative to SEQ ID NO: 2.
67. The nucleic acid of claim 66, wherein the at least one amino acid substitution is a substitution selected from the group consisting of G1057D, S931R, A289P, V996S, L1100F, M287A, C482S, M685L, S547P, V916I, A1256E, M1259A, N332G, M358L, Y919P, T950A, M376P, P151K, D584L, Y1162Q, M640S, C314L, R1131G, A965S,

Y1044L, D1236G, A970L, Q1039R, C474Y, and T404, or combinations thereof relative to SEQ ID NO: 2.

68. The nucleic acid of claim 66, wherein the at least one amino acid substitution is a substitution in at least one of positions S931, Q1039 or M287 relative to SEQ ID NO: 2.
69. The nucleic acid of any one of claims 1-46 and 48-68, wherein the R2 protein or portion thereof comprises at least one amino acid deletion, wherein the at least one amino acid deletion comprises a deletion of a N-terminal intrinsically disordered region or a portion thereof of the R2 protein.
70. The nucleic acid of claim 69, wherein the at least one amino acid deletion is a deletion of between about 10 amino acids and about 300 amino acids within the N-terminal intrinsically disordered region.
71. The nucleic acid of claim 69, wherein the N-terminal intrinsically disordered region is located at about 10 amino acids and about 300 amino acids within the N-terminus of the WT R2 protein.
72. The nucleic acid of any one of claims 69-71, wherein the at least one amino acid deletion is a deletion of between about 10 amino acids and about 300 amino acids within the N-terminus of the R2 protein relative to SEQ ID NO: 1.
73. The nucleic acid of any one of claims 69-71, wherein the at least one amino acid deletion is a deletion of between about 160 amino acids and about 240 amino acids at the N-terminus of the R2 protein relative to SEQ ID NO: 1.
74. The nucleic acid of any one of claims 69-71, wherein the at least one amino acid deletion is a deletion of about 161 amino acids, about 184 amino acids, or about 218 amino acids, at the N-terminus the R2 protein relative to SEQ ID NO: 1.
75. The nucleic acid of any one of claims 1-45 and 48-74, wherein the R2 protein or portion thereof comprises a native reverse transcriptase domain and does not comprise a functional native endonuclease domain.
76. The nucleic acid of claim 75, wherein the at least one heterologous polypeptide comprises a heterologous endonuclease domain.
77. The nucleic acid of claim 76, wherein the heterologous endonuclease domain is an endonuclease domain of another R2 protein.

78. The nucleic acid of claim 77, wherein the another R2 protein has an amino acid sequence set forth in any one of SEQ ID NOS: 1, 2, 400-480 and 30001-30060, or an amino acid sequence that is at least 70% identical thereto.
79. The nucleic acid of any one of claims 1-45 and 48-74, wherein the R2 protein comprises a native endonuclease domain and does not comprise a functional native reverse transcriptase domain.
80. The nucleic acid of claim 79, wherein the at least one heterologous polypeptide comprises the heterologous reverse transcriptase domain.
81. The nucleic acid of claim 80, wherein the heterologous reverse transcriptase domain is a reverse transcriptase domain of another WT R2 protein.
82. The nucleic acid of claim 81, wherein the another R2 protein has an amino acid sequence set forth in any one of SEQ ID NOS: 1, 2, 400-480 and 30001-30060, or an amino acid sequence that is at least 70% identical thereto.
83. The nucleic acid of claim 80, wherein heterologous reverse transcriptase domain comprises an amino acid sequence set forth in SEQ ID NO:295, or an amino acid sequence that is at least 70% identical thereto.
84. The nucleic acid of any one of claims 1-83, wherein the nucleic acid is a RNA molecule, a DNA molecule or a hybrid RNA/DNA molecule.
85. The nucleic acid of any one of claims 1-84, wherein the nucleic acid comprises one or more chemical or sequence modifications.
86. The nucleic acid of claim 85, wherein the one or more chemical or sequence modifications are selected from the group consisting of an RNA CAP, a modified polyA length, a linker, a chemical modification, a 5' UTR modification, a 3' UTR modification, a modified Kozak sequence, a modified (e.g., truncated) stem loop, an RNA stabilization motif, a 5-methoxyuridine (5-MO-U) modification, a 5-methylcytidine (5mC) modification, and one or more additional and/or modified microsatellites.
87. The nucleic acid of any one of claims 1-86, wherein the engineered driver comprises a heterologous CTIP polypeptide and a heterologous HMGN1 polypeptide.
88. The nucleic acid of claim 87, wherein the heterologous CTIP polypeptide and the heterologous HMGN1 polypeptide are situated on the N-terminal side of the R2 protein or portion thereof.

89. The nucleic acid of claim 88, wherein the heterologous CTIP polypeptide is situated on the N-terminal side of the heterologous HMGN1 polypeptide.
90. The nucleic acid of any one of claims 87-89, wherein the R2 protein or portion thereof comprises a deletion within a native N-terminal intrinsically disordered region.
91. The nucleic acid of any one of claims 87-90, wherein the engineered driver comprises a heterologous intrinsically disordered region.
92. The nucleic acid of claim 91, wherein the heterologous intrinsically disordered region is situated on the N-terminal side of the heterologous CTIP polypeptide and the heterologous HMGN1 polypeptide.
93. The nucleic acid of claim 92, wherein the engineered driver has an amino acid sequence set forth in any one of SEQ ID NOS: 679-714, or an amino acid sequence that is at least 70% identical thereto.
94. The nucleic acid of any one of claims 87-90, wherein the engineered driver comprises at least one heterologous RNA/DNA processing polypeptide.
95. The nucleic acid of claim 94, wherein the at least one heterologous RNA/DNA processing polypeptide comprises a RNA helicase polypeptide.
96. The nucleic acid of claim 95, wherein the RNA helicase polypeptide is situated at the N-terminal side of the heterologous CTIP polypeptide and the heterologous HMGN1 polypeptide.
97. The nucleic acid of claim 96, wherein the engineered driver has an amino acid sequence set forth in SEQ ID NO: 783, or an amino acid sequence that is at least 70% identical thereto.
98. The nucleic acid of any one of claims 1-86, wherein the engineered driver comprises a heterologous Sto7d polypeptide and a heterologous HMGN1 polypeptide.
99. The nucleic acid of claim 98, wherein the heterologous Sto7d polypeptide and the heterologous HMGN1 polypeptide are situated on the N-terminal side of the R2 protein or portion thereof.
100. The nucleic acid of claim 99, wherein the heterologous Sto7d polypeptide is situated on the N-terminal side of the heterologous HMGN1 polypeptide.
101. The nucleic acid of any one of claims 98-100, wherein the R2 protein or portion thereof comprises a deletion within a native N-terminal intrinsically disordered region.

102. The nucleic acid of any one of claims 98-101, wherein the engineered driver comprises a RNA/DNA processing polypeptide.
103. The nucleic acid of any one of claims 102, wherein the RNA/DNA processing polypeptide comprises a heterologous RNA helicase recruitment motif polypeptide.
104. The nucleic acid of claim 103, wherein the heterologous RNA helicase recruitment motif polypeptide is situated on the N-terminal side of the heterologous Sto7d polypeptide and the heterologous HMGN1 polypeptide.
105. The nucleic acid of claim 104, wherein the heterologous RNA helicase recruitment motif polypeptide is a TUDOR polypeptide.
106. The nucleic acid of claim 105, wherein the engineered driver comprises an amino acid sequence set forth in SEQ ID NO: 1234, or an amino acid sequence that is at least 70% identical thereto.
107. A nucleic acid encoding an engineered driver comprising an R2 protein or portion thereof, a heterologous CTIP polypeptide, and a heterologous HMGN1 polypeptide.
108. The nucleic acid of claim 107, wherein the heterologous CTIP polypeptide and the heterologous HMGN1 polypeptide are situated on the N-terminal side of the R2 protein or portion thereof.
109. The nucleic acid of claim 108, wherein the heterologous CTIP polypeptide is situated on the N-terminal side of the heterologous HMGN1 polypeptide.
110. The nucleic acid of any one of claims 107-109, wherein the R2 protein or portion thereof comprises a deletion in a N-terminal intrinsically disordered region or portion thereof.
111. The nucleic acid of any one of claims 107-110, wherein the engineered driver comprises a heterologous intrinsically disordered region.
112. The nucleic acid of claim 111, wherein the heterologous intrinsically disordered region is situated on the N-terminal side of the heterologous CTIP polypeptide and the heterologous HMGN1 polypeptide.
113. The nucleic acid of claim 112, wherein the engineered driver comprises, in a N-terminal to C-terminal direction, the heterologous intrinsically disordered region, the heterologous CTIP polypeptide, the heterologous HMGN1 polypeptide, and the R2 protein or portion thereof.

114. The nucleic acid of claim 113, wherein the engineered driver comprises an amino acid sequence set forth in SEQ ID NOS: 679-714, or an amino acid sequence that is at least 70% identical thereto.
115. The nucleic acid of any one of claims 107-110, wherein the engineered driver further comprises a heterologous RNA helicase polypeptide.
116. The nucleic acid of claim 115, wherein the engineered driver comprises, in a N-terminal to C-terminal direction, the heterologous RNA helicase polypeptide, the heterologous CTIP polypeptide, the heterologous HMGN1 polypeptide, and the R2 protein or portion thereof.
117. The nucleic acid of claim 116, wherein the engineered driver comprises an amino acid sequence set forth in SEQ ID NO: 783, or an amino acid sequence that is at least 70% identical thereto.
118. A nucleic acid encoding an engineered driver comprising an R2 protein or portion thereof, a heterologous Sto7d polypeptide, and a heterologous HMGN1 polypeptide.
119. The nucleic acid of claim 118, wherein the heterologous Sto7d polypeptide and the heterologous HMGN1 polypeptide are situated on the N-terminal side of the R2 protein or portion thereof.
120. The nucleic acid of claim 119, wherein the heterologous Sto7d polypeptide is situated on the N-terminal side of the heterologous HMGN1 polypeptide.
121. The nucleic acid of any one of claims 118-120, wherein the R2 protein or portion thereof comprises a deletion in a N-terminal intrinsically disordered region or portion thereof.
122. The nucleic acid of any one of claims 118-121, comprising a heterologous TUDOR polypeptide.
123. The nucleic acid of claim 122, wherein the engineered driver comprises, in a N-terminal to C-terminal direction, the heterologous Tudor polypeptide, the heterologous CTIP polypeptide, the heterologous HMGN1 polypeptide, and the R2 protein or portion thereof.
124. The nucleic acid of claim 123, wherein the engineered driver comprises an amino acid sequence set forth in SEQ ID NO: 1234, or an amino acid sequence that is at least 70% identical thereto.

125. A nucleic acid encoding an engineered driver comprising an amino acid sequence with one or more modifications relative to any one of SEQ ID NOS: 30001-30060.
126. The nucleic acid of claim 125, wherein the engineered driver comprises an amino acid sequence set forth in SEQ ID NOS: 30001-30060, or an amino acid sequence that is at least 70% identical thereto.
127. The nucleic acid of claim 125 or claim 126, wherein the engineered driver comprises a reverse transcriptase domain and an endonuclease domain.
128. The nucleic acid of any one of claims 125-127, wherein the one or more modifications is at least one amino acid substitution, at least one amino acid deletion, at least one amino acid addition, or a combination thereof.
129. The nucleic acid of any one of claims 125-127, wherein the engineered driver comprises an amino acid sequence with one or more modifications relative to a sequence selected from the group consisting of SEQ ID NOS: 30015, 30019, 30020, 30027, and 30029.
130. A nucleic acid encoding an engineered driver comprising a R2 protein or portion thereof, wherein the R2 protein or portion thereof comprises the amino acid substitution of D555K relative to SEQ ID NO:1.
131. The nucleic acid of claim 130, wherein the R2 protein or portion thereof comprises the amino acid substitutions P280K, D555K, T1011S, I1219N, and A1282G relative to SEQ ID NO:1.
132. The nucleic acid of claim 130, wherein the R2 protein or portion thereof comprises the amino acid substitutions G257R, P280K, N363H, E399M, V539G, H423Q, A442P, D537Y, K547R, D555K, R669L, N889G, F973E, T1011A, Q1019R, D1133P, K1085R, R1177G, Y1242K, and A1282G relative to SEQ ID NO:1.
133. A nucleic acid encoding an engineered driver comprising an RTE polypeptide fused to a heterologous intrinsically disordered polypeptide.
134. The nucleic acid of claim 133, wherein the heterologous intrinsically disordered domain polypeptide comprises an amino acid sequence set forth in SEQ ID NO: 1171-1216 or an amino acid sequence that is at least 70% identical thereto.
135. The nucleic acid of claim 133, wherein the RTE polypeptide is lacking a native intrinsically disordered region or portion thereof.

136. A nucleic acid encoding an engineered driver comprising an RTE polypeptide fused to a RNA helicase polypeptide.
137. The nucleic acid of claim 136, wherein the RNA helicase polypeptide is a EIF4A polypeptide.
138. The nucleic acid of claim 137, wherein the EIF4A polypeptide comprises an amino acid sequence set forth in SEQ ID NO: 585, or an amino acid sequence that is at least 70% identical thereto.
139. The nucleic acid of claim 136, wherein the RNA helicase polypeptide is a Dengue virus RNA helicase polypeptide.
140. The nucleic acid of claim 139, wherein the Dengue virus RNA helicase polypeptide comprises an amino acid sequence set forth in SEQ ID NO: 1875, or an amino acid sequence that is at least 70% identical thereto.
141. An engineered driver encoded by the nucleic acid of any one of claims 1-140.
142. A transgene expression system comprising:
- a. a driver construct comprising the nucleic acid of any one claims 1-140, or the engineered driver of claim 141; and
 - b. a gene delivery construct comprising a transgene and at least one retroelement untranslated region (RTE-UTR) derived from a RTE, wherein the at least one RTE-UTR is capable of binding the engineered driver, and the engineered driver is configured to mediate integration of the transgene into a genome.
143. A transgene expression system comprising:
- a. a driver construct comprising the nucleic acid of any one claims 1-132, or the engineered driver encoded by the nucleic acid of any one of claim 1-132; and
 - b. a gene delivery construct comprising a transgene and at least one retroelement untranslated region (RTE-UTR) derived from a R2 RTE, wherein the at least one RTE-UTR is capable of binding the engineered driver, and the engineered driver is configured to mediate integration of the transgene into a genome.
144. The system of claim 143, wherein the driver construct and the gene delivery construct are in a trans configuration and provided as separate nucleic acid constructs.
145. The system of claim 143, wherein the driver construct and the gene delivery construct are in a cis configuration and provided in the same nucleic acid construct.

146. The system of any one of claims 143-145, wherein the at least one RTE-UTR comprises a 3' RTE-UTR, a 5' RTE-UTR, or a combination thereof.
147. The system of claim 146, wherein the 3' RTE-UTR has a nucleic acid sequence set forth in any one of SEQ ID NOS: 215, 217, 219, 20754, and 33001-33060.
148. The system of claim 146, wherein the 5' RTE-UTR has a nucleic acid sequence set forth in any one of SEQ ID NOS: 216, 218, 220, 20755, and 35001-35060.
149. The system of claim 146, wherein the 3' RTE-UTR has a nucleic acid sequence set forth in any one of SEQ ID NOS: 1338-1340, 1344-1347, 1352-1354, 1843, and 1865.
150. The system of claim 146 or claim 147 or claim 149, wherein the 5' RTE-UTR has a nucleic acid sequence set forth in any one of SEQ ID NOS: 1341-1343, 1348-1351, 1355-1360-1365, 1366, 1842, and 1866.
151. The system of claim 146 or claim 147 or claim 149, wherein the 5' RTE-UTR has a nucleic acid sequence set forth in any one of SEQ ID NOS: 1838-1850, 1873-1874, 1862-1863, 1867-1872.
152. The system of any one of claims 142-151, wherein the R2 protein or portion thereof encoded by the driver construct is derived from the same R2 RTE as the at least one RTE-UTR comprised in the gene delivery construct.
153. The system of any one of claims 142-151, wherein the R2 protein or portion thereof encoded by the driver construct is derived from a heterologous R2 RTE with respect to the at least one RTE-UTE comprised in the gene delivery construct.
154. The system of any one of claims 146-148, wherein the 3' RTE-UTR has a nucleic acid sequence set forth in SEQ ID NO: 33027, and the 5' RTE-UTR has a sequence set forth in SEQ ID NO: 35027
155. The system of any one of claims 146-148, wherein the 3' RTE-UTR has a nucleic acid sequence set forth in SEQ ID NO: 33019, and the 5' RTE-UTR has a sequence set forth in SEQ ID NO: 35019
156. The system of claim 154 or claim 155, wherein the R2 protein or portion thereof comprised in the engineered driver has an amino acid sequence set forth in SEQ ID NO: 1, or an amino acid sequence including at least one amino acid substitution, at least one amino acid deletion, at least one amino acid addition, or a combination thereof, relative to SEQ ID NO: 1.

157. The system of claim 156, wherein the engineered driver comprises an amino acid sequence set forth in SEQ ID NO: 636 or SEQ ID NO:1396, or an amino acid sequence that is at least 70% identical thereto.
158. The system of any one of claims 142-157, wherein the transgene encodes a gene selected from the group consisting of: Factor VIII, Factor IX, Phenylalanine hydroxylase, ATP7B, alpha glucosidase, argininosuccinate synthetase, galactose-1-phosphate uridylyltransferase, and ornithine transcarbamylase.
159. The system of any one of claims 142-158, wherein the transgene comprises an intron.
160. The system of any one of claims 142-159, wherein the transgene is in a sense orientation.
161. The system of any one of claims 142-159, wherein the transgene is in an antisense orientation.
162. The system of any one of claims 142-161, wherein the gene delivery construct is a RNA molecule, a DNA molecule or a hybrid RNA/DNA molecule.
163. A gene delivery construct comprising:
- a. a transgene; and
 - b. at least one R2 retroelement untranslated region (RTE-UTR) derived from a R2 RTE, wherein the RTE-UTR is capable of binding a R2 protein, and wherein:
 - the at least one RTE-UTR comprises a 3' RTE-UTR, a 5' RTE-UTR, or a combination thereof,
 - the 3' RTE-UTR has a nucleic acid sequence set forth in any one of SEQ ID NOS: 1338-1340, 1344-1347, 1352-1354, 1843, 1865, and 33001-33060, and
 - the 5' RTE-UTR has a nucleic acid sequence set forth in any one of SEQ ID NOS: 1341-1343, 1348-1351, 1355-1360-1365, 1366, 1838-1850, 1873-1874, 1862-1863, 1866-1872, and 35001-35060
164. The gene delivery construct of claim 163, wherein the 3' RTE-UTR has a nucleic acid sequence set forth in SEQ ID NO: 33027, and the 5' RTE-UTR has a nucleic acid sequence set forth in SEQ ID NO: 35027.
165. The gene delivery construct of claim 163, wherein the 3' RTE-UTR has a nucleic acid sequence set forth in SEQ ID NO: 33019, and the 5' RTE-UTR has a nucleic acid sequence set forth in SEQ ID NO: 35019

166. The gene delivery construct of claim 163 or claim 164, wherein the gene delivery construct is a RNA molecule, a DNA molecule or a hybrid RNA/DNA molecule.
167. The gene delivery construct of any one of claims 163-166, wherein the transgene encodes a gene selected from the group consisting of: Factor VIII, Factor IX, Phenylalanine hydroxylase, ATP7B, alpha glucosidase, argininosuccinate synthetase, galactose-1-phosphate uridylyltransferase, and ornithine transcarbamylase.
168. The gene delivery construct of any one of claims 163-167, wherein the transgene comprises an intron.
169. The gene delivery construct of any one of claims 163-168, wherein the transgene is in a sense orientation.
170. The gene delivery construct of any one of claims 163-169, wherein the transgene is in an antisense orientation.
171. The gene delivery construct of any one of claims 163-170, wherein the gene delivery construct is a RNA molecule, a DNA molecule or a hybrid RNA/DNA molecule.
172. The gene delivery construct of any one of claims 163-166, wherein the transgene encodes a gene selected from the group consisting of: Factor VIII, Factor IX, Phenylalanine hydroxylase, ATP7B, alpha glucosidase, argininosuccinate synthetase, galactose-1-phosphate uridylyltransferase, and ornithine transcarbamylase.
173. A lipid nanoparticle (LNP) comprising:
- a. the nucleic acid of any one of claims 1-140;
 - b. the engineered driver of claim 141;
 - c. the transgene expression system of any one of claims 142-162; or
 - d. the gene delivery construct of any one of claim 163-172; or
 - e. a combination thereof.
174. The LNP of claim 173, wherein the LNP comprises an ionizable lipid, a PEGylated lipid, an amino lipid, a phospholipid, a cholesterol, a sterol, a non-cationic lipid, or a combination thereof.
175. A pharmaceutical composition comprising:
- a. the nucleic acid of any one of claims 1-140;
 - b. the engineered driver of claim 141;
 - c. the transgene expression system of any one of claims 142-162;
 - d. the gene delivery construct of any one of claim 163-172;

- e. the LNP of any one of claims 173-174; or
- f. a combination thereof;

and a pharmaceutically acceptable carrier.

176. A method of promoting integration of a transgene into a genome of a cell, the method comprising contacting a cell with:
- a. the nucleic acid of any one of claims 1-140;
 - b. the engineered driver of claim 141;
 - c. the transgene expression system of any one of claims 142-162;
 - d. the gene delivery construct of any one of claim 163-172;
 - e. the LNP of any one of claims 173-174;
 - f. a composition of claim 175; or
 - g. a combination thereof.
177. The method of claim 176, wherein the cell is a quiescent cell.
178. The method of claim 176, wherein the cell is an actively dividing cell.
179. The method of claim 176, wherein the cell is T-cell.
180. The method of claim 176, wherein the cell is a cancer cell.
181. The method of any one of claims 176-177 and 180, wherein the cell is an endothelial cell, an astrocyte, an osteoblast, a lung cell, a neuron, a hepatocyte, a pancreatic cell, a myocyte, a keratinocyte, a bladder cell, or an adipocyte.
182. The method of any one of claims 176-181, wherein the method takes place *ex vivo*.
183. The method of any one of claims 176-181, wherein the cell is an allogeneic cell.
184. The method of any one of claims 176-181, wherein the cell is an autologous cell.
185. The method of any one of claims 176-184, wherein the method takes place *ex vivo*.
186. The method of any one of claims 176-184, wherein the method takes place *in vivo*.
187. A population of cells prepared by the method of any one of claims 176-186.
188. A method of treating a disease in a subject in need thereof, comprising administering to the subject a therapeutically effective amount of:
- a. the nucleic acid of any one of claims 1-140;

- b. the engineered driver of claim 141;
- c. the transgene expression system of any one of claims 142-162;
- d. the gene delivery construct of any one of claim 163-172;
- e. the LNP of any one of claims 173-174;
- f. the composition of claim 175; or
- g. a combination thereof.

189. The nucleic acid of any one of claims 1-140, the engineered driver of claim 141, the transgene expression system of any one of claims 142-162, the gene delivery construct of any one of claim 163-172, the LNP of any one of claims 173-174, the composition of claim 175, or a combination thereof, for use in a method of treating the human or animal body by therapy.

190. Use of the nucleic acid of any one of claims 1-140, the engineered driver of claim 141, the transgene expression system of any one of claims 142-162, the gene delivery construct of any one of claim 163-172, the LNP of any one of claims 173-174, the composition of claim 175, or a combination thereof, for the manufacture of a medicament.