# *IN VIVO* MODIFICATION OF CELL GENOMES

## CROSS-REFERENCE TO RELATED APPLICATIONS

**[0001]** This application claims the benefit of priority to U.S. Provisional Application No. 63/556,286, filed February 21, 2024, and U.S. Provisional Application No. 63/717,732, filed November 7, 2024, the contents of which are incorporated herein by reference in their entirety.

## BACKGROUND

**[0002]** CD19-targeting CAR T cells have shown remarkable results B-ALL, but many aspects of this technology remain to be improved (June, C. H. & Sadelain, M. Chimeric Antigen Receptor Therapy. N. Engl. J. Med. **379**, 64–73 (2018)). Relapses occur in a large proportion of patients, and clinical results remain poor in solid tumors, in part due limited CAR T cell persistence and T cell dysfunction. Wide implementation of T cell therapy is limited to using randomly integrating vectors in autologous patient cells, which involves high cost and variability of the final product. CRISPR-Cas9 and AAV-mediated homology-directed repair (HDR) were previously used to achieve targeted integration of CAR transgenes into the endogenous human TCR α locus (termed *TRAC*-CARs). These methods resulted in optimal and dynamic regulation of CAR expression, prevented T cell exhaustion, and resulted in superior control of NALM6 xenograft tumors model (Eyquem, J. et al. Targeting a CAR to the *TRAC* locus with CRISPR-Cas9 enhances tumor rejection. Nature 543, 113–117 (2017)). This work has been useful for the development of allogeneic CAR T cells therapy because the methods simultaneously disrupt the T cell receptor (TCR), which can attenuate of graft-versus-host disease (GVHD). Clinical trials using allogeneic *TRAC*-CAR T cells generated from healthy donor or induced pluripotent stem cells (iPSCs) derived T cells have achieved complete responses in patients with hematological malignancies when associated with deep lymphodepleting pre-conditioning (Poirot, L. et al. Multiplex Genome-Edited T-cell Manufacturing Platform for 'Off-the-Shelf' Adoptive T-cell Immunotherapies. Cancer Res. 75, 3853–3864 (2015); Qasim, W. et al. Molecular remission of infant B-ALL after infusion of universal TALEN gene-edited CAR T cells. Sci. Transl. Med. 9, eaaj2013 (2017); van

der Stegen, S. J. C. et al. Generation of T-cell-receptor-negative CD8αβ-positive CAR T cells from T-cell-derived induced pluripotent stem cells. Nat. Biomed. Eng. 6, 1284–1297 (2022)).

**[0003]** However, while certain manufacturing hurdles are addressed, allogeneic CAR T cells can be eventually rejected by the host immune system, and relapses have been observed. *In vivo* generation of CAR T cells is an exciting new development that would move the field from cell to gene therapy which has several potential advantages, including bypassing leukapheresis and costly manufacturing while using less differentiated T cells. As of January 2024, CAR T cells have been generated *in vivo* using randomly integrating virus vectors or lipid nanoparticles (LNPs) resulting in transient CAR expression (Pfeiffer, A. et al. *In vivo* generation of human CD19-CAR T cells results in B-cell depletion and signs of cytokine release syndrome. EMBO Mol. Med. 10, e9158 (2018); Smith, T. T. et al. *In situ* programming of leukemia-specific T cells using synthetic DNA nanocarriers. Nat. Nanotechnol. 12, 813–820 (2017)). Such methods have their own set of challenges, including the need to design efficient gene delivery systems and the potential for off-target effects. Delivery and CAR expression can be T cell specific as off-target engineering of hematopoietic stem cells (HSCs) could lead to transformational mutagenesis and CAR expression in tumor cells could prevent cell surface expression of the CAR target and cause antigen negative relapse (Booth, C., Gaspar, H. B. & Thrasher, A. J. Treating Immunodeficiency through HSC Gene Therapy. Trends Mol. Med. 22, 317–327 (2016); Ruella, M. et al. Induction of resistance to chimeric antigen receptor T cell therapy by transduction of a single leukemic B cell. Nat. Med. 24, 1499–1503 (2018)).

## SUMMARY

**[0004]** Provided herein are methods and compositions related to *in vivo*, *ex vivo* or *in vitro* modification a gene from or in a mammal.

**[0005]** In one aspect, a method of in vivo DNA insertion into a gene in cells in a mammal comprises administering to the mammal: (a) a first vector that delivers a polynucleotide-guided nuclease to a cell nuclei, and (b) a second vector comprising a donor template polynucleotide. In some embodiments, a guide polynucleotide for the polynucleotide-guided nuclease is present in the first vector, the second vector, or both the first and second vectors. In some embodiments, the guide polynucleotide is targeted to a portion of a gene in the cell genome that is selectively expressed in the cells, wherein the gene is any gene in the cell genome. In some embodiments, the

presence of the first vector and the second vector in the cells in the mammal results in cleavage of the gene that is selectively expressed in the cells and integration of the donor template polynucleotide into the gene. In some embodiments, the first vector and/or the second vector selectively targets T cells. In some embodiments, the first vector comprises the polynucleotide-guided nuclease or a polynucleotide encoding polynucleotide-guided nuclease. In another aspect, a method of DNA insertion into a gene in mammalian cells comprising contacting to the cells (for example but not limited to, administering to the mammal): (a) a first vector that delivers a polynucleotide-guided nuclease to a cell nuclei, and (b) a second vector comprising a donor template polynucleotide.

**[0006]** In some embodiments, the first vector is an enveloped delivery vehicle (EDV), wherein the EDV optionally comprises a cell-specific binding molecule.

**[0007]** In some embodiments, the first vector is a lipid nanoparticle (LNP), wherein the LNP optionally comprises a cell-specific binding molecule.

**[0008]** In some embodiments, the first vector and/or the second vector selectively targets one or more type of immune cells.

**[0009]** In some embodiments, the immune cells are T cells and the cell-specific binding molecule binds to CD3, CD4, CD5, CD7, CD8, CD28, 4-1BB ligand, T cell receptor (TCR) α constant chain, TCR β constant chain, or a major histocompatibility complex (MHC) carrying T cell receptor (TCR) specific peptide.

**[0010]** In some embodiments, the immune cells are B cells and the cell-specific binding molecule binds to CD19, CD20, BCMA, CD138, TACI, or CD22.

**[0011]** In some embodiments, the immune cells are NK cells and the cell-specific binding molecule binds to CD56, CD16, NKp46/NCR1, NCR2, or KIR.

**[0012]** In some embodiments, the immune cells are monocytes or macrophages and the cell-specific binding molecule binds to CD11b, CD68, CD14, CD33, or CD163.

**[0013]** In some embodiments, the immune cells are dendritic cells and the cell-specific binding molecule binds to CD11b, CD11c, XCR1, CD33, CD1c, or CD123.

**[0014]** In some embodiments, the first vector and/or the second vector selectively targets stem cells. In some embodiments, the stem cells are hematopoietic stem cells (HSCs) and the cell-specific binding molecule binds to CD34, CD117, CD49f, CD38, CD90, or EPCR.

**[0015]** In some embodiments, the first vector further comprises a protein that catalyzes membrane fusion. In some embodiments, the first vector comprises the polynucleotide encoding the polynucleotide-guided nuclease and wherein the polynucleotide is an RNA molecule.

**[0016]** In some embodiments, the second vector is a recombinant Adeno-associated virus (AAV) vector that has T-cell tropism. In some embodiments, the AAV vector comprises an AAV capsid variant having reduced antibody-mediated neutralization compared to a wild-type capsid.

**[0017]** In some embodiments, the guide polynucleotide is a guide RNA and the polynucleotide-guided nuclease is a CRISPR-Cas endonuclease.

**[0018]** In some embodiments, the gene in the T cell genome is T cell receptor α constant (TRAC), T cell receptor β constant (TRBC), T cell receptor γ constant (TRGC), T cell receptor δ constant (TRDC), CD4, CD5, CD6, CD7, CD8a, CD8b, CD3e, CD247, CD27, CD28, interleukin-2R (IL-2R) α, IL-2R beta, killer cell lectin-like receptor (KLR) C1 (KLRC1), KLRF1, KLRG1, granzyme (GZM) A (GZMA), GZMB, GZMH, GZMK, Zap-70, lymphocyte-specific protein tyrosine kinase (LCK), linker for activation of T cells (LAT), IL-2 inducible T cell kinase (ITK), transcription factor 7 (TCF7), lymphoid enhancer binding factor 1 (LEF1), or FoxP3.

**[0019]** In some embodiments, the guide polynucleotide targets the polynucleotide-guided nuclease to an exon of a TRAC gene. In some embodiments, the guide polynucleotide targets the polynucleotide-guided nuclease to exon 1 of the TRAC gene. In some embodiments, the guide polynucleotide targets the polynucleotide-guided nuclease to an intron of a TRAC gene.

**[0020]** In some embodiments, the gene in the B cell genome is CD19, CD20, CD22, CD138, BCMA, TACI, MS4A1, IGH, IGK, CD79A, or CD79B.

**[0021]** In some embodiments, the gene in the NK cell genome is NCAM1, FCGR3A, NCR1, NCR2, KLRC1, NKG2D, NKG7, KIR2DL1, KIR2DL2, KIR2DL3, KIR2DL4, KIR3DL1 or KIR3DL1.

**[0022]** In some embodiments, the gene in the monocyte or macrophage genome is CD11b, CD11c, CD14, CD33, CD163, CLEC7A, C1QA, C1QB, C1QC, or MSR1.

**[0023]** In some embodiments, the gene in the dendritic cell genome is CD1C, DCIR, CLEC10A, NDRG2, or TPM2.

**[0024]** In some embodiments, the gene in the HSC genome is PTPRC, CD34, HBB, or RAG2.

**[0025]** In some embodiments, the protein that catalyzes membrane fusion is a vesicular stomatitis virus glycoprotein G protein (VSVG) or a fusogenic variant thereof.

**[0026]** In some embodiments, (a) the first vector is (i) an EDV optionally comprising a T cell-specific binding molecule and a protein that promotes membrane fusion, or (ii) an LNP optionally comprising a T cell-specific binding molecule; and (b) the second vector is a recombinant Adeno-associated virus (AAV) vector having T cell tropism. In some embodiments, the guide polynucleotide targets the polynucleotide-guided nuclease to exon 1 of a TRAC gene.

**[0027]** In some embodiments, the first vector is (a) the EDV, wherein the EDV comprises the polynucleotide-guided nuclease, or (b) the LNP, wherein the LNP comprises a polynucleotide encoding the polynucleotide-guided nuclease and wherein the polynucleotide is an RNA molecule.

**[0028]** In some embodiments, the donor template polynucleotide comprises a coding sequence for a polypeptide comprising a chimeric antigen receptor (CAR), a T cell receptor (TCR), or an HLA-independent T cell receptor (HIT).

**[0029]** In some embodiments, the donor polynucleotide comprises a coding sequence for a polypeptide comprising (a) an extracellular target-binding domain, (b) a transmembrane domain, (c) a hinge domain, and (d) an intracellular signaling domain. In some embodiments, the polypeptide extracellular target-binding domain binds to a CD19 polypeptide. In some embodiments, the polypeptide transmembrane domain and the polypeptide hinge domain are derived from a CD28 polypeptide. In some embodiments, the polypeptide intracellular signaling domain comprises a modified CD3ζ polypeptide comprising (a) a native ITAM1, (b) a modified ITAM2 comprising two Tyr to Phe mutations, (c) a modified ITAM3 comprising two Tyr to Phe mutations, (d) a native BRS1, (e) a native BRS2, (f) a native BRS3, and (g) a co-stimulatory signaling region comprising a CD28 polypeptide.

**[0030]** In some embodiments, the donor template polynucleotide is a homology-dependent repair template (HDRT) polynucleotide, homology-mediated end-joining template (HMEJT) polynucleotide, or a homology-independent targeted integration template (HITIT) polynucleotide.

**[0031]** In some embodiments, after integration of the donor template polynucleotide into the gene, the coding sequence from the donor template polynucleotide is under control of endogenous promoter and/enhancer sequences.

**[0032]** In another aspect, a composition comprises (a) a first vector that delivers a polynucleotide-guided nuclease to a cell nuclei, and (b) a second vector comprising a donor template polynucleotide, and (c) a pharmaceutical carrier and/or a pharmaceutical excipient. In some embodiments, a guide polynucleotide for the polynucleotide-guided nuclease is present in the first vector, the second vector, or both the first and second vectors; and the guide polynucleotide is targeted to a portion of a gene in the cell genome that is selectively expressed in the cells.

**[0033]** In some embodiments, the first vector is an EDV or an LNP, wherein the EDV or the LNP optionally comprises a cell-specific binding molecule.

**[0034]** In some embodiments, the cell-specific binding molecule is a T cell-specific binding molecule that binds to CD3, CD4, CD5, CD7, CD8, CD28, 4-1BB ligand, T cell receptor (TCR) α constant chain, TCR β constant chain, or a major histocompatibility complex (MHC) carrying T cell receptor (TCR) specific peptide.

**[0035]** In some embodiments, the cell-specific binding molecule is a B cell-specific binding molecule that binds to to CD19, CD20, BCMA, CD138, TACI, or CD22.

**[0036]** In some embodiments, the cell-specific binding molecule is an NK cell-specific binding molecule that binds to CD56, CD16, NKp46/NCR1, NCR2, or KIR.

**[0037]** In some embodiments, the cell-specific binding molecule is a monocyte or macrophage-specific binding molecule that binds to CD11b, CD68, CD14, CD33, or CD163.

**[0038]** In some embodiments, the cell-specific binding molecule is a dendritic cell-specific binding molecule that binds to CD11b, CD11c, XCR1, CD33, CD1c, or CD123.

**[0039]** In some embodiments, the cell-specific binding molecule is an HSC-specific binding molecule that binds to CD34, CD117, CD49f, CD38, CD90, or EPCR.

**[0040]** In some embodiments, wherein the first vector further comprises a protein that catalyzes membrane fusion.

**[0041]** In some embodiments, the second vector is a recombinant Adeno-associated virus (AAV) vector that has T-cell tropism. In some embodiments, the AAV vector comprises an AAV capsid variant having reduced antibody-mediated neutralization compared to a wild-type capsid.

**[0042]** In some embodiments, the guide polynucleotide is a guide RNA and the polynucleotide-guided nuclease is a CRISPR-Cas endonuclease. In some embodiments, the guide polynucleotide targets the polynucleotide-guided nuclease to an exon of a TRAC gene. In some embodiments, the guide polynucleotide targets the polynucleotide-guided nuclease to exon 1 of the TRAC gene.

**[0043]** In some embodiments, the protein that catalyzes membrane fusion is a vesicular stomatitis virus glycoprotein G protein (VSVG) or a fusogenic variant thereof.

**[0044]** In some embodiments, the donor template polynucleotide comprises a coding sequence for a polypeptide comprising a chimeric antigen receptor (CAR), a T cell receptor (TCR), or an HLA-independent T cell receptor (HIT).

**[0045]** In some embodiments, the donor template polynucleotide comprises a coding sequence for a polypeptide comprising (a) an extracellular target-binding domain, (b) a transmembrane domain, (c) a hinge domain, and (d) an intracellular signaling domain. In some embodiments, the polypeptide extracellular target-binding domain binds to a CD19 polypeptide. In some embodiments, the polypeptide transmembrane domain and the polypeptide hinge domain are derived from a CD28 polypeptide. In some embodiments, the polypeptide intracellular signaling domain comprises a modified CD3ζ polypeptide comprising (a) a native ITAM1, (b) a modified ITAM2 comprising two Tyr to Phe mutations, (c) a modified ITAM3 comprising two Tyr to Phe mutations, (d) a native BRS1, (e) a native BRS2, (f) a native BRS3, and (g) a co-stimulatory signaling region comprising a CD28 polypeptide.

**[0046]** In some embodiments, the donor template polynucleotide is a homology-dependent repair template (HDRT) polynucleotide, homology-mediated end-joining template (HMEJT) polynucleotide, or a homology-independent targeted integration template (HITIT) polynucleotide.

## BRIEF DESCRIPTION OF THE DRAWINGS

**[0047]** Figure 1A-1D show delivery of homology-directed templates (HDR) with Adeno-associated virus vectors (AAVs) and delivery of Cas9 with enveloped delivery vehicles (EDVs) for knock-in at the T cell receptor α constant (TRAC) locus. Figure 1A shows an illustration of an HDR template targeting a chimeric antigen receptor (CAR) to TRAC delivered by an AAV and an EDV carrying Cas9/sgTRAC. Figure 1B shows TCR expression in human T cells after transduction with EDVs carrying Cas9/sgTRAC ribonucleoproteins (RNPs). TCR expression was determined by flow cytometry. 50 µl of EDV was used for $3\times10^5$ T cells (Low), $2\times10^5$ T cells (Med), $1\times10^5$ T cells (High) and compared to untransduced T cells (UT). Figure 1C shows enhanced green fluorescent protein (EGFR) and T cell receptor (TCR) expression after transduction with Cas9/sgTRAC RNP EDV alone (EDV), Cas9/sgTRAC RNP EDV and AAV6 (EDV + AAV6), or Cas9/sgTRAC RNP EDV and Ark312 (EDV + Ark312), as determined by flow. Figure 1D shows knock-in efficiency determined by EGFR expression of activated T cells transduced with EDV, EDV + AAV6, or EDV + Ark312, each at three different multiplicities of infection (MOIs).

**[0048]** Figure 2A-2D shows characterization of human T cells after transduction with EDVs carrying Cas9/sgTRAC RNP and VSVG-WT (VSVG-WT EDV), or EDVs Cas9/sgTRAC RNP and VSVGm-aCD3 (VSVGm-aCD3 EDV), in combination with AAV6 or Ark312. Figure 2A shows TCR and EGFR expression in human T cells after transduction with different combinations of Cas9/sgTRAC RNP EDV with VSVG-WT (VSVG-WT EDV), Cas9/sgTRAC RNP EDV with VSVGm-aCD3 (VSVGm-aCD3 EDV), AAV6 and Ark312. EGFR and TCR expression were determined by flow cytometry. Figure 2B shows knock-in efficiency of a 1928z-1XX CAR at TRAC by combining either VSVG-WT EDV or VSVGmut-aCD3 EDV carrying Cas9/sgTRAC, with either AAV6 or Ark312 carrying an HDRT. Figure 2C shows CD25 expression as determined by flow cytometry as a marker for T cell activation. UT = untreated cells; Low = cells treated with 5 µl concentrated EDVs; High = cells treated with 25 µl concentrated EDVs. Figure 2D compares cytotoxicity of TCR knock out (KO) T cells with TRAC-1928z-1XX CAR T cells generated by transduction with VSVG-WT EDVs carrying Cas9/sgTRAC with either AAV6 or Ark312 (312) carrying a HDRT. Cytotoxicity was determined by luminescence. Results are the mean ± SEM from three technical replicates.

**[0049]** Figure 3A-3D show treatment of NOD scid gamma (NSD) mice and flow cytometry analysis of spleen cells isolated from the mice. Figure 3A shows a treatment schedule for NSD mice engrafted with human peripheral blood mononuclear cells (PBMCs). Mice received intravenous (IV) injections of Cas9/sgTRAC EDVs with VSVG-WT or VSVGm-aCD3 (aCD3), in combination with AAV6 or Ark312 carrying HDRT targeting a 1928z-1XX-P2A-EGFRT construct to TRAC. Figure 3B shows total B cells and CAR-T cells per spleen. Figure 3C shows CD45+/CD19+ B cells and CD45+/CAR+/EGFRT+ CAR-T cells as a percentage of CD45+ cells in each spleen. Figure 3D shows total number CD4+/CD45+/CAR+/EGFRT+ CAR-T cells and CD8+/CD45+/CAR+/EGFRT+ CAR-T cells for each spleen. Figures 3B-3D: Results are the mean ± SEM. Significance was assessed using a repeated-measures one-way ANOVA and Dunnett's multiple comparisons test. *$p<0.05$; **$p<0.01$.

**[0050]** **Figure 4A-C: 4a**, Illustration of an EDV expressing a WT VSVG protein (VSVG-WT, yellow) and an EDV expressing a mutated VSVG protein and an anti-CD3 scFv (αCD3, red). **4b**, Cells were treated with VSVG-WT or αCD3-EDVs carrying Cas9 and a sgRNA targeting *CLTA*, and AAV6 or Ark312 carrying a HDR template to knock-in a reporter sfGFP transcript in exon 1 of *CLTA*. GFP expression is only observed if correctly integrated into *CLTA* exon 1. Cells were treated with $3\times10^5$ sgRNA/cell EDV and $5\times10^5$ vg/cell AAV. GFP expression was analysed by flow cytometry at least 72 hours after treatment and genomic integration was confirmed by dPCR. **4c**, Primary human T cells (n=3), NK cells (n=2), $CD34^+$ HSC (n=3), Macrophages (n=3) and four cell lines from B cell malignancies (NALM6, Raji, SupB15 JeKo) were treated according to (**4b**). All cell types were positive for GFP signal when treated with VSVG-WT + AAV6, and the relative knock-in rate (heat map below) was calculated by normalizing each condition to the VSVG-WT + AAV6 values (plotted above).

**[0051]** **Figure 5A-D**: **5a**, Cells were treated with VSVG-WT or αCD3-EDVs carrying Cas9 and a sgRNA targeting *CLTA*, and AAV6 or Ark312 carrying a HDR template to knock-in a reporter sfGFP transcript in exon 1 of *CLTA*. GFP expression is only observed if correctly integrated in *CLTA*. Cells were treated with $3\times10^5$ sgRNA/cell EDV, and with $5\times10^5$ vg/cell AAV. GFP expression was analysed by flow cytometry at least 72 hours after treatment and genomic integration was confirmed by dPCR. Primary human T cells (n=3), NK cells (n=2), $CD34^+$ HSC (n=3), Macrophages (n=3). Results are the mean ± SEM from two or three donors (n=2 or 3). **5b**

Four cell lines from B cell malignancies NALM6, Raji, SupB15 JeKo were treated according to **5a**. GFP expression was analyzed by flow cytometry at least 72 hours after treatment. **5c**, Human $CD34^+$ hematopoietic stem cells were electroporated with Cas9/sg*CLTA* and treated with AAV6 or Ark312 carrying a HDR template to knock-in a reporter sfGFP transcript in exon 1 of *CLTA*. Cells were treated $5\times10^4$ vg/cell AAV. GFP expression was analysed by flow cytometry at least 72 hours after treatment (left panel) and genomic integration was confirmed by dPCR (right panel). Results are the mean ± SEM from three donors (n=3). **5d**, dPCR analysis on genomic DNA extracted from samples in (**5a**) to confirm genomic integration at *CLTA*. Including knock-in analysis of treated NALM6-ffLuc-GFP cells treated the same way as **5a** (flow cytometry analysis not available due to ubiquitous GFP expression). Data presented as percent of edited alleles. Results are the mean ± SEM from three donors (n=3).

**[0052]** **Figure 6A-F. 6a**, Schematic of tumour challenge model. NSG-MHCI/II dKO mice were injected with $2.5\times10^5$ NALM6-ffLuc-GFP cells intravenously, followed by an intraperitoneal injection of $2\times10^7$ human PBMCs three days later. One day following the PBMC injections, the mice were injected with either PBS (PBMC only) or αCD3-EDV ($2.5\times10^{11}$ sgRNA per mouse) and Ark312 ($1\times10^{12}$ vg per mouse) carrying an HDR template targeting 1928z-1XX CAR and EGFRT to *TRAC*. Tumour burden was tracked by bioluminescence imaging (BLI) measurements. At day 39 post EDV/AAV injection, the mice were rechallenged with $5\times10^6$ NALM6-ffLuc-GFP cells intravenously and euthanized for organ harvest after 13 days. **6b**, BLI measurements in mice injected with NALM6 (n=5), NALM6 and PBMC (n=10), or NALM6, PBMC and EDV/AAV (n=9). Rechallenge with NALM6 was performed in five EDV/AAV treated mice that controlled tumour growth by day 39, and in five age-matched control mice. BLI values are the average of dorsal and ventral signals and presented as photons/s/$cm^2$. 6**c**, Kaplan-Meier survival analysis until day 39 comparing mice injected with NALM6 only (n=5), NALM6 and PBMC (n=10), or NALM6, PBMC and EDV/AAV (n=9). Significance was assessed using a log-rank (Mantel-Cox) test. **6d**, Total number of CAR-T cells in bone marrow and spleen in NALM6 rechallenged mice from (**6a-b**). CAR expression was determined by flow cytometry. Results are the mean ± SEM from five mice (n=5). **6e**, Representative flow plots showing TCR and CAR expression in human $CD45^+$ cells from bone marrow and spleen from (**6d**). **6f**, CD4 and CD8 expression in CAR-T cells from bone marrow and spleen in NALM6 rechallenged mice from (**6a-b**) as determined by flow

cytometry. Results are the mean ± SEM from five mice (n=5). **6c**, Significance was assessed using a log-rank (Mantel-Cox) test. ***p<0.001.

**[0053]** **FIG. 7A-C. 7a**, CAR-T cell percentages in human $CD45^+$ cells from bone marrow and spleen in NALM6 rechallenged mice from (**Fig. 6a-b**). CD45 and CAR expressions were determined by flow cytometry. Results are the mean ± SEM from six mice (n=5). **7b**, NALM6 ($CD19^+/GFP^+/CD45^-$) percentages in bone marrow and spleen in NALM6 rechallenged mice from (**Fig. 6a-b**). CD45, GFP and CD19 expressions were determined by flow cytometry. Results are the mean ± SEM from six mice (n=5). **7c**, Total number of NALM6 cells ($CD19^+/GFP^+/CD45^-$) bone marrow and spleen in NALM6 rechallenged mice from (**Fig. 6a-b**). CD45, GFP and CD19 expressions were determined by flow cytometry. Results are the mean ± SEM from six mice (n=5).

**[0054]** Figures 8A-8C show that Ark315 enables efficient *in vivo* T cell engineering. Figure 8A: Schematic showing NSG-MHCI/II dKO mice transplanted with human PBMC were injected with CD3-EDV ($2.5 \times 10^{11}$ sgRNA per mouse) carrying Cas9/sgTRAC and either Ark312 or Ark315 ($1 \times 10^{12}$ vg per mouse) carrying an HDR template targeting 1928z-1XX CAR and EGFRT to TRAC. Spleens were harvested and isolated for flow cytometry 14 days after EDV/AAV injections. Figure 8B: Representative flow plot showing EGFRt and CAR expression in splenic T cells from Figure 8A. Figure 8C: Percentage of CAR positive and CD19 positive in human CD45+ cells were determined by flow 938 cytometry in spleens from mice in Figure 8A.

**[0055]** Figures 9A-9C show that combining EDV and AAV can generate TRAC-TCR T cells. Figure 9A: Schematic representation of a mutant VSVG CD3 retargeted EDV (top portion) carrying TRAC Cas9-RNP and an Ark312 AAV (bottom portion) carrying an HDR template to knock-in an NY-ESO TCR transgene at the TRAC locus. Figure 9B: Activated human T cells were treated with EDVs carrying Cas9/sgTRAC ( $3 \times 10^5$ MOI based on sgRNA quantification) and TRAC-NYESO Ark312 ($2 \times 10^5$ MOI). Figure 9C: Average TRAC-TCR KI of 3 different human PBMC donors engineered with CD3-EDV and TRAC-NYESO-TCR AAV.

**[0056]** Figures 10A-10B show that combining LNP and AAV can generate TRAC CAR T cells. Figure 10A: Schematic representation of an carrying Cas9 mRNAand TRAC sgRNA packaged in a LNP and an AAV6 carrying an HDR template to knock-in a CAR and EGFRt transgene at the TRAC locus. Activated human T cells were co-transduced with the AAV6 and LNP and analyzed at day 5. Figure 10B: Representative flow plot showing EGFRt and CAR expression.

**[0057]** Figure 11 shows that engineered AAV variants mediate TRAC-specific CAR knock-in in human T cells. Activated primary human T cells were simultaneously transduced with VSVG-EDVs packaging TRAC-targeted Cas9 RNPs and an increasing MOI of various engineered AAVs encoding an anti-CD19-1XX CAR (19-1XX) and EGFRt between human (hu) TRAC homology arms. HDR rate represents the fraction of cells with a TCR disruption (TCRab-) that also expressed EGFRt. EDV only represents cells transduced with VSVG-EDVs but no AAV. Values are the mean of two biological donors. Error bars represent the standard deviation. AAV-031, AAV-027, and AAV-026 represent three independent AAV productions of the same capsid and transgene.

## TERMINOLOGY

**[0058]** For the purposes of promoting an understanding of the principles of the present disclosure, reference will now be made to preferred embodiments and specific language will be used to describe the same. It will nevertheless be understood that no limitation of the scope of the disclosure is thereby intended, such alteration and further modifications of the disclosure as illustrated herein, being contemplated as would normally occur to one skilled in the art to which the disclosure relates.

**[0059]** As used herein, the term "vector" refers to a vehicle for introducing a molecule, e.g., nucleic acids, proteins, small molecules, or combinations thereof, into a target cell. In some embodiments, the nucleic acids, proteins, small molecules, or combinations thereof are packaged within the vector. In some embodiments, a vector is also used for the integration of a nucleotide sequence (e.g., a gene or a coding sequence) into a target nucleic acid of the target cell, and optionally, for the expression of the nucleotide sequence (e.g., a gene or a coding sequence) by that cell. Vector forms include, for example, plasmids or virus vectors, as well polymer particles, enveloped delivery vehicles (EDVs), virus-like particles (VLPs), and lipid nanoparticles (LNPs).

**[0060]** As used herein, the term "enveloped delivery vehicle," "EDV," "virus like particle," and "VLP" are synonymous and refer to a vector comprising a group-specific antigen (gag) protein fragment and a lipid membrane. EDVs may be used to delivery one or more cargo molecules to a cell. In some cases, an EDV also comprises a targeting polypeptide that provides for specific binding to a target cell, thereby facilitating delivery of its cargo a particular target cell. In some cases, EDVs are used to encapsidate or otherwise carry a cargo protein, for example, a polynucleotide-guided nuclease. In instances where the cargo protein is a polynucleotide-guided

nuclease, the EDV may also encapsidate a guide polynucleotide. In some cases, EDVs are used to encapsidate nucleic acids, for example, a nucleic acid with a coding sequence for a protein of interest, e.g., a polynucleotide-guided nuclease. Examples of EDVs can be found in Hamilton, Jennifer R et al. "Targeted delivery of CRISPR-Cas9 and transgenes enables complex immune cell engineering." Cell reports vol. 35,9 (2021): 109207. doi:10.1016/j.celrep.2021.109207; Hamilton, Jennifer R et al. "In vivo human T cell engineering with enveloped delivery vehicles." Nature biotechnology, 10.1038/s41587-023-02085-z. 11 Jan. 2024, doi:10.1038/s41587-023-02085-z; and U.S. Patent Application Publication No. 2022/0403379.

**[0061]** As used herein, the terms "lipid particle," "lipid nanoparticle," and "LNP" refer to a particle comprising lipids and may comprise phospholipids and/or ionizable lipids. An LNP may comprise additional lipid components, such as a sterol and/or a conjugated lipid, and may further comprise a nucleic acid, wherein the nucleic acid may be encapsulated within the LNP. In some embodiments, the LNP also comprises a polynucleotide-guided nuclease and/or a guide polynucleotide within the LNP.

**[0062]** In cases where the vector is a virus vector, the virus (e.g., AAV) particle functions as a molecule (e.g., nucleic acid) delivery vehicle, and encapsidates the vector genome (e.g., viral DNA or vDNA) within the virus particle (i.e., the virion). Alternatively, in some contexts, the term "vector" can be used to refer to the virus vector genome/vDNA alone, and the vector is capable of expressing all components required for encapsidating its cargo molecules and their delivery of to a target cell.

**[0063]** The genomic sequences of various serotypes of AAV and the autonomous parvoviruses, as well as the sequences of the native terminal repeats (TRs), Rep proteins, and capsid subunits are known in the art. Non-limiting examples of AAV genomic sequences include GenBank Accession Numbers NC_002077, NC_001401, NC_001729, NC_001863, NC_001829, NC_001862, NC_000883, NC_001701, NC_001510, NC_006152, NC_006261, AF063497, U89790, AF043303, AF028705, AF028704, J02275, J01901, J02275, X01457, AF288061, AH009962, AY028226, AY028223, NC_001358, NC_001540, AF513851, AF513852, and AY530579.

**[0064]** A "rAAV vector genome" or "rAAV genome" as used herein is an AAV genome (i.e., vDNA) that comprises one or more heterologous nucleic acid sequences. rAAV vectors generally

require only the terminal repeat(s) (TR(s)) in cis to generate virus. All other viral sequences are dispensable and can be supplied in trans (Muzyczka, (1992) Curr. Topics Microbiol. Immunol. 158:97). Typically, the rAAV vector genome will only retain the one or more TR sequences so as to maximize the size of the transgene that can be efficiently packaged by the vector. The structural and non-structural protein coding sequences can be provided in trans (e.g., from a vector, such as a plasmid, or by stably integrating the sequences into a packaging cell). In some embodiments, a disclosed rAAV vector genome comprises at least one TR sequence (e.g., AAV TR sequence), optionally two TRs (e.g., two AAV TRs), which typically will be at the 5′ and 3′ ends of the vector genome and flank the heterologous nucleic acid, but need not be contiguous thereto. The TRs can be the same or different from each other.

**[0065]** The term "terminal repeat" or "TR" includes any viral terminal repeat or synthetic sequence that forms a hairpin structure and functions as an inverted terminal repeat (i.e., mediates the desired functions such as replication, virus packaging, integration and/or provirus rescue, and the like). The TR can be an AAV TR or a non-AAV TR. For example, a non-AAV TR sequence such as those of other parvoviruses (e.g., canine parvovirus (CPV), mouse parvovirus (MVM), human parvovirus B-19) or any other suitable virus sequence (e.g., the SV40 hairpin that serves as the origin of SV40 replication) can be used as a TR, which can further be modified by truncation, substitution, deletion, insertion and/or addition. Further, the TR can be partially or completely synthetic, such as the "double-D sequence" as described in U.S. Pat. No. 5,478,745 to Samulski et al.

**[0066]** An "AAV terminal repeat" or "AAV TR" can be from any AAV, including but not limited to serotypes 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13 or any other AAV now known or later discovered. An AAV terminal repeat need not have the native terminal repeat sequence (e.g., a native AAV TR sequence can be altered by insertion, deletion, truncation and/or missense mutations), as long as the terminal repeat mediates the desired functions, e.g., replication, virus packaging, integration, and/or provirus rescue, and the like.

**[0067]** An AAV vector typically comprises a protein-based capsid, and a nucleic acid encapsidated by the capsid. The nucleic acid can be, for example, a vector genome comprising a transgene flanked by inverted terminal repeats. The AAV "capsid" is a near-spherical protein shell that comprises individual "capsid proteins" or "subunits." AAV capsids typically comprise about

60 capsid protein subunits, associated and arranged with T=1 icosahedral symmetry. When an AAV vector is described herein as comprising an AAV capsid protein, it will be understood that the AAV vector comprises a capsid, wherein the capsid comprises one or more AAV capsid proteins (i.e., subunits). Also described herein are "viral-like particles" or "virus-like particles," which refers to a capsid that does not comprise any vector genome or nucleic acid comprising a transgene.

**[0068]** The virus vectors of the present disclosure can further be "targeted" virus vectors (e.g., having a directed tropism) and/or a "hybrid" parvovirus (i.e., in which the viral TRs and viral capsid are from different parvoviruses) as described in International Patent Application Publication WO 2000/028004 and Chao et al. (2000) Molecular Therapy. 2:619.

**[0069]** The term "self-complimentary AAV" or "scAAV" refers to a recombinant AAV vector which forms a dimeric inverted repeat DNA molecule that spontaneously anneals, resulting in earlier and more robust transgene expression compared with conventional single-strand (ss) AAV genomes. See, e.g., McCarty, D.M., et al., Gene Therapy 8, 1248- 1254 (2001). Unlike conventional ssAAV, scAAV can bypass second-strand synthesis, the rate-limiting step for gene expression. Moreover, double-stranded scAAV is less prone to DNA degradation after viral transduction, thereby increasing the number of copies of stable episomes. Notably, scAAV can typically only hold a genome that is about 2.4 kb, half the size of a conventional AAV vector. In some embodiments, the AAV vectors described herein are self-complementary AAVs.

**[0070]** As used herein, the term "adeno-associated virus" (AAV), includes but is not limited to, AAV type 1, AAV type 2, AAV type 3 (including types 3A and 3B), AAV type 4, AAV type 5, AAV type 6, AAV type 7, AAV type 8, AAV type 9, AAV type 10, AAV type 11, AAV type 12, AAV type 13, AAV type rh32.33, AAV type rh8, AAV type rh10, avian AAV, bovine AAV, canine AAV, equine AAV, ovine AAV, and any other AAV now known or later discovered. See, e.g., BERNARD N. FIELDS et al., VIROLOGY, volume 2, chapter 69 (4th ed., Lippincott-Raven Publishers). A number of AAV serotypes and clades have been identified (see, e.g., Gao et al, (2004) J. Virology 78:6381 -6388; Moris et al, (2004) Virology 33-:375-383; and Table 1).

**[0071]** As used herein, the term "homology-directed repair" or "HDR" refers to a cellular process in which cut or nicked ends of a DNA strand are repaired by polymerization from a homologous template nucleic acid. Thus, the original sequence is replaced with a sequence on the

"HDR template" or "HDRT." In some embodiments, an insertion sequence, for example, a DNA template, can be introduced to obtain a specific HDR-induced change of the sequence at a target site. In this way, specific mutations can be introduced at a cut site, for example, a cut site created by a targeted nuclease. A single-stranded insertion sequence or a double-stranded insertion sequence can be used by a cell as a template for editing or modifying the genome of a cell, for example, by HDR. Generally, the single-stranded insertion sequence or the double-stranded insertion sequence has a pair of homology sequences, with one on each side of the insertion sequence ("flanking homology region" or "flanking homology sequence") to be inserted at a target cut or insertion site; the homology sequences have sufficient complementarity to a target site for HDR to occur. In some cases, each flanking homology region is at least about 20, 30, 40, 50, 100, 150, 200, 250, 300, 350, 400 or 450 nucleotides. In some cases, a nucleotide sequence that is homologous to a genomic sequence is at least 80%, 90%, 95%, 99% or 100% complementary to the genomic sequence.

**[0072]** As used herein, the term "homology-mediated end-joining" or "HMEJ" refers to a process for integrating exogenous DNA fragments into the genome using a nuclease and a linearized donor polynucleotide, i.e., a HMEJ template (HMEHT) polynucleotide. The targeted genomic locus and the donor polynucleotide are simultaneously cleaved by a nuclease (e.g., CRISPR-Cas nuclease) and then connected to each other, resulting in transgene integration. In some cases, the donor polynucleotide harbors at least one guide polynucleotide (e.g., gRNA) target site. In some cases, the donor polynucleotide has homology arms that are each at least 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950 or nucleotides long, with sequences that are complementary to sequences at the targeted genomic locus. In some cases, a nucleotide sequence that is homologous to a genomic sequence is at least 80%, 90%, 95%, 99% or 100% complementary to the genomic sequence.

**[0073]** As used herein, the term "homology-independent targeted integration" or "HITI" refers to a cellular process in which the DNA double-stranded breaks are repaired using non-homologous end-joining (NHEJ). While in some cases, the DNA-double stranded break strands can be ligated back together without a template, in other cases, the DNA-double stranded break strands are repaired with a HITI template (HITIT) polynucleotide. The HITIT polynucleotide contains an insertion sequence and does not have homology with the target site. Thus, the original sequence at

a target site is replaced with the insertion sequence of the HITIT polynucleotide. In some cases, HITI is used in conjunction with a polynucleotide-guided nuclease, e.g., a CRISPR-Cas nuclease. The nuclease simultaneously creates double-stranded breaks in both the target sequence and the HITIT polynucleotide, thereby generating blunt ends. Thus, in some cases, the HITIT polynucleotide also contains a nuclease cleavage site at either one or both ends of the HITIT polynucleotide insertion sequence.

**[0074]** The term "introducing," as used in the context of molecules, e.g., nucleic acids and/or proteins, of the present disclosure, refers to presenting one or more molecules to a target cell in such a manner that the molecules may gain access to the interior of the target cell.

**[0075]** As used herein, the terms "modifying," "editing," and the like, used in the context of modifying a genome of a cell, refers to inducing a structural change in the sequence of the genome at a target genomic region. As used herein, the terms "gene modification," "gene editing," and the like refer to transient or permanent genetic modification in a target cell following introduction of molecules, e.g., nucleic acids and/or proteins, of the present disclosure, into the cell. The genetic modification can be accomplished by contacting a target cell with one or more vectors that can deliver the nucleic acids and/or proteins for introducing the genetic modification, and inserting the nucleic acids into the genome of the cell. The choice of method is generally dependent on the type of cell being transformed and the circumstances under which the transformation is taking place (i.e. *in vitro*, *ex vivo*, or *in vivo*). In some embodiments, the gene modification is performed, for example, by inducing a double stranded break within a target genomic region, or a pair of single stranded nicks on opposite strands and flanking the target genomic region. Methods for inducing single or double stranded breaks at or within a target genomic region include the use of a polynucleotide-guided nuclease (e.g., transcription activator-like effector nuclease (TALEN), a zinc finger nuclease, a meganuclease, Cas9 nuclease domain, or derivative thereof), and a guide polynucleotide (e.g., a guide RNA or pair of guide RNAs) directed to the target genomic region.

**[0076]** As used herein, the terms "chimeric antigen receptor" and "CAR" refer to artificial multi-module molecules capable of triggering or inhibiting the activation of an immune cell which generally but not exclusively comprise an extracellular domain (e.g., a ligand/antigen binding domain), a transmembrane domain and one or more intracellular signaling domains. The term CAR is not limited specifically to CAR molecules but also includes CAR variants. CAR variants include

split CARs wherein the extracellular portion (e.g., the ligand binding portion) and the intracellular portion (e.g., the intracellular signaling portion) of a CAR are present on two separate molecules. CAR variants also include ON-switch CARs which are conditionally activatable CARs, e.g., comprising a split CAR wherein conditional hetero-dimerization of the two portions of the split CAR is pharmacologically controlled. CAR variants also include bispecific CARs, which include a secondary CAR binding domain that can either amplify or inhibit the activity of a primary CAR. CAR variants also include inhibitory chimeric antigen receptors (iCARs) which may, e.g., be used as a component of a bispecific CAR system, where binding of a secondary CAR binding domain results in inhibition of primary CAR activation. CAR molecules and derivatives thereof (i.e., CAR variants) are described, e.g., in PCT Application No. US2014/016527; Fedorov et al. Sci Transl Med (2013) ;5(215):215ra172; Glienke et al. Front Pharmacol (2015) 6:21; Kakarla & Gottschalk 52 Cancer J (2014) 20(2):151-5; Riddell et al. Cancer J (2014) 20(2):141-4; Pegram et al. Cancer J (2014) 20(2):127-33; Cheadle et al. Immunol Rev (2014) 257(1):91-106; Barrett et al. Annu Rev Med (2014) 65:333-47; Sadelain et al. Cancer Discov (2013) 3(4):388-98; Cartellieri et al., J Biomed Biotechnol (2010) 956304.

**[0077]** The terms "polynucleotide" and "nucleic acid" are used interchangeably to refer to a polymeric form of nucleotides of any length, either deoxyribonucleotides or ribonucleotides. The terms include RNA, DNA, and synthetic forms and mixed polymers of the above. In particular embodiments, a nucleotide refers to a ribonucleotide, deoxynucleotide or a modified form or analog of either type of nucleotide (e.g., an RNA molecule may have an artificial cap or may comprise one or more pseudouridine or other non-natural nucleoside), and combinations thereof. A reference to a nucleic acid sequence encompasses its complement unless otherwise specified. Thus, a reference to a nucleic acid molecule having a particular sequence should be understood to encompass its complementary strand, with its complementary sequence. Reference to a "polynucleotide" or "nucleic acid" that encodes a polypeptide sequence also includes codon-optimized nucleic acids and nucleic acids that comprise alternative codons that encode the same polypeptide sequence.

**[0078]** As used herein, the term "complementary" or "complementarity" refers to specific base pairing between nucleotides or nucleic acids. Complementary nucleotides are, generally, A and T (or A and U), and G and C. The guide RNAs described herein can comprise sequences, for

example, DNA targeting sequences that are perfectly complementary or substantially complementary (e.g., having 1-4 mismatches) to a genomic sequence.

**[0079]** The term "gene" can refer to the segment of DNA involved in producing or encoding a polypeptide chain. It may include regions preceding and following the coding region (leader and trailer) as well as intervening sequences (introns) between individual coding segments (exons). Alternatively, the term "gene" can refer to the segment of DNA involved in producing or encoding a non-translated RNA, such as an rRNA, tRNA, guide RNA (*e.g.*, a single guide RNA), or micro RNA.

**[0080]** A "promoter" is defined as one or more a nucleic acid control sequences that direct transcription of a nucleic acid. As used herein, a promoter includes nucleic acid sequences near the start site of transcription, such as, in the case of a polymerase II type promoter, a TATA element. A promoter also optionally includes distal enhancer or repressor elements, which can be located as much as several thousand base pairs from the start site of transcription.

**[0081]** "Polypeptide," "peptide," and "protein" are used interchangeably herein to refer to a polymer of amino acid residues. As used herein, the terms encompass amino acid chains of any length, including full-length proteins, wherein the amino acid residues are linked by covalent peptide bonds. The terms can also refer to genetically coded and non-genetically coded amino acids, chemically or biochemically modified or derivatized amino acids, and polypeptides having modified peptide backbones. The term includes fusion proteins, including, but not limited to, fusion proteins with a heterologous amino acid sequence, fusions with heterologous and homologous sequences, with or without N-terminal methionine residues; immunologically tagged proteins; and the like.

**[0082]** The phrase "percent identical," "percent identity," or equivalents used in the context of two nucleic acids or polypeptides, refers to a sequence that has at least a specified level of identity, *e.g.*, at least 50% sequence identity with a reference sequence (*e.g.*, any SEQ ID NO included herein). Alternatively, percent identity can be any integer from 50% to 100%. Some embodiments include at least: 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99%, compared to a reference sequence using the programs described herein, *e.g.*, BLAST using standard parameters, as described below.

**[0083]** For sequence comparison, typically one sequence acts as a reference sequence, to which test sequences are compared. When using a sequence comparison algorithm, test and reference sequences are entered into a computer, subsequence coordinates are designated, if necessary, and sequence algorithm program parameters are designated. Default program parameters can be used, or alternative parameters can be designated. The sequence comparison algorithm then calculates the percent sequence identities for the test sequences relative to the reference sequence, based on the program parameters.

**[0084]** A "comparison window," as used herein, includes reference to a segment of any one of the number of contiguous positions selected from the group consisting of from 20 to 600, usually about 50 to about 200, more usually about 100 to about 150 in which a sequence may be compared to a reference sequence of the same number of contiguous positions after the two sequences are optimally aligned. Methods of alignment of sequences for comparison are well-known in the art. Optimal alignment of sequences for comparison can be conducted, *e.g.,* by the local homology algorithm of Smith & Waterman, *Adv. Appl. Math.* 2:482 (1981), by the homology alignment algorithm of Needleman & Wunsch, *J. Mol. Biol.* 48:443 (1970), by the search for similarity method of Pearson & Lipman, *Proc. Nat'l. Acad. Sci. USA* 85:2444 (1988), by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI), or by manual alignment and visual inspection.

**[0085]** Algorithms that are suitable for determining percent sequence identity and sequence similarity are the BLAST and BLAST 2.0 algorithms, which are described in Altschul *et al.* (1990) *J. Mol. Biol.* 215: 403-410 and Altschul *et al.* (1977) *Nucleic Acids Res.* 25: 3389-3402, respectively. Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information (NCBI) web site. The algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul *et al, supra*). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are then extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are

calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always >0) and N (penalty score for mismatching residues; always <0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences) uses as defaults a word size (W) of 28, an expectation (E) of 10, M=1, N=-2, and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a word size (W) of 3, an expectation (E) of 10, and the BLOSUM62 scoring matrix (*see* Henikoff & Henikoff, *Proc. Natl. Acad. Sci. USA* 89:10915 (1989)).

**[0086]** The BLAST algorithm also performs a statistical analysis of the similarity between two sequences (*see, e.g.*, Karlin & Altschul, *Proc. Nat'l. Acad. Sci. USA* 90:5873-5787 (1993)). One measure of similarity provided by the BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a nucleic acid is considered similar to a reference sequence if the smallest sum probability in a comparison of the test nucleic acid to the reference nucleic acid is less than about 0.01, more preferably less than about $10^{-5}$, and most preferably less than about $10^{-20}$.

**[0087]** For sequence comparison, typically one sequence acts as a reference sequence, to which test sequences are compared. When using a sequence comparison algorithm, test and reference sequences are entered into a computer, subsequence coordinates are designated, if necessary, and sequence algorithm program parameters are designated. Default program parameters can be used, or alternative parameters can be designated. The sequence comparison algorithm then calculates the percent sequence identities for the test sequences relative to the reference sequence, based on the program parameters.

**[0088]** As used herein, the term "endogenous" with reference to a nucleic acid, for example, a gene, or a protein in a cell is a nucleic acid or protein that occurs in that particular cell as it is found

in nature, for example, at its natural genomic location or locus. Moreover, a cell "endogenously expressing" a nucleic acid or protein expresses that nucleic acid or protein as it is found in nature.

**[0089]** As used herein the phrase "heterologous" refers to what is not found in nature. The term "heterologous sequence" refers to a sequence not normally found in a given cell in nature, i.e., not wild-type (WT). As such, a heterologous nucleotide or protein sequence may be: (a) foreign to its host cell (*i.e.,* is exogenous to the cell); (b) naturally found in the host cell (*i.e.,* endogenous) but present at an unnatural quantity in the cell (*i.e.,* greater or lesser quantity than naturally found in the host cell); or (c) be naturally found in the host cell but positioned outside of its natural locus. A heterologous nucleotide or polypeptide sequence can also mean a sequence that is not found in the native nucleic acid or protein, respectively. For example, relative to a T cell receptor, a chimeric antigen receptor (CAR) comprises an amino acid sequence with portions that are each derived a protein other than the T cell receptor.

**[0090]** The term "recombinant" or "engineered" when used with reference, *e.g.*, to a cell, or nucleic acid, protein, or vector, indicates that the cell, nucleic acid, protein, or vector, has been modified by the introduction of a heterologous nucleic acid or protein or the alteration of a native nucleic acid or protein, or that the cell is derived from a cell so modified. Such modifications are often accomplished by manipulation of isolated segments of nucleic acids and may include, for example, genetic engineering techniques. Thus, for example, recombinant cells or engineered cells express genes that are not found within the native (non-recombinant or non-engineered) form of the cell, or the recombinant cells or engineered cells express native genes that are otherwise abnormally expressed, under expressed or not expressed at all.

**[0091]** As used herein, the term "specifically binds" to a target refers to a binding reaction whereby, for example, a targeting polypeptide binds to the target with greater affinity, greater avidity, and/or greater duration than it binds to a different target. In some embodiments, a target-binding protein has at least 2-fold, 3-fold, 4-fold, 5-fold, 6-fold, 7-fold, 8-fold, 9-fold, 10-fold, 20-fold, 25-fold, 50-fold, 100-fold, 1,000-fold, 10,000-fold, or greater affinity for the target compared to an unrelated target when assayed under the same binding affinity assay conditions. The term "specific binding," "specifically binds to," or "is specific for" a particular target, as used herein, can be exhibited, for example, by a molecule (*e.g.,* a ligand) having an equilibrium dissociation constant, $K_D$, for the target of, *e.g.,* $10^{-2}$ M or smaller, *e.g.,* $10^{-3}$ M, $10^{-4}$ M, $10^{-5}$ M, $10^{-6}$ M, $10^{-7}$ M,

$10^{-8}$ M, $10^{-9}$ M, $10^{-10}$ M, $10^{-11}$ M, or $10^{-12}$ M. In some embodiments, a target-binding domain has a $K_D$ of less than 100 nM or less than 10 nM.

**[0092]** As used herein, the term "hematopoietic stem cell" or "HSC" refers to a type of stem cell that can give rise to a blood cell. HSCs can give rise to cells of the myeloid or lymphoid lineages, or a combination thereof. HSCs are predominantly found in the bone marrow, although they can be isolated from peripheral blood, or a fraction thereof. Various cell surface markers can be used to identify, sort, or purify hematopoietic stem cells. In some cases, HSCs are identified as c-kit$^+$ and lin$^-$. In some cases, human HSCs are identified as CD34$^+$, CD59$^+$, Thy1/CD90$^+$, CD38$^{lo/-}$, C-kit/CD117$^+$, lin$^-$. In some cases, human HSCs are identified as CD34$^-$, CD59$^+$, Thy1/CD90$^+$, CD38$^{lo/-}$, C-kit/CD117$^+$, lin$^-$. In some cases, human HSCs are identified as CD133$^+$, CD59$^+$, Thy1/CD90$^+$, CD38$^{lo/-}$, C-kit/CD117$^+$, lin$^-$. In some cases, the HSCs are CD150$^+$CD48$^-$CD244$^-$.

**[0093]** As used herein, the phrase "hematopoietic cell" refers to a cell derived from a hematopoietic stem cell. The hematopoietic cell may be obtained or provided by isolation from an organism, system, organ, or tissue (*e.g.*, blood, or a fraction thereof). Alternatively, an hematopoietic stem cell can be isolated and the hematopoietic cell obtained or provided by differentiating the stem cell. Hematopoietic cells include cells with limited potential to differentiate into further cell types. Such hematopoietic cells include, but are not limited to, multipotent progenitor cells, lineage-restricted progenitor cells, common myeloid progenitor cells, granulocyte-macrophage progenitor cells, or megakaryocyte-erythroid progenitor cells. Hematopoietic cells include cells of the lymphoid and myeloid lineages, such as lymphocytes, erythrocytes, granulocytes, monocytes, and thrombocytes. In some embodiments, the hematopoietic cell is an immune cell, such as a T cell, B cell, macrophage, a natural killer (NK) cell or dendritic cell. In some embodiments the cell is an innate immune cell.

**[0094]** As used herein, the term "immune cells" generally includes white blood cells (leukocytes) which are derived from HSCs produced in the bone marrow. Examples of immune cells include lymphoid cells (T cells, B cells, and natural killer (NK) cells) and myeloid cells (neutrophil, eosinophil, basophil, monocyte, macrophage, and dendritic cells).

**[0095]** As used herein, the term "T cell" refers to a lymphoid cell that expresses a T cell receptor molecule. T cells include, but are not limited to, naïve T cells, stimulated T cells, primary T cells (*e.g.*, uncultured), cultured T cells, immortalized T cells, helper T cells, cytotoxic T cells, memory

T cells, regulatory T cells, natural killer T cells, combinations thereof, or sub-populations thereof. T cells can be $CD4^+$, $CD8^+$, or $CD4^+$ and $CD8^+$. T cells can be helper cells, for example helper cells of type $T_h1$, $T_h2$, $T_h3$, $T_h9$, $T_h17$, or $T_{FH}$. T cells can be cytotoxic T cells. Regulatory T cells can be $FOXP3^+$ or $FOXP3^-$. T cells can be α/Beta T cells or gamma/delta T cells. In some embodiments, the T cell is a $CD4^+CD25^{hi}CD127^{lo}$ regulatory T cell. In some embodiments, the T cell is a regulatory T cell selected from the group consisting of Tr1, Th3, CD8+CD28-, Treg17, and Qa-1 restricted T cells, or a combination or sub-population thereof. In some embodiments, the T cell is a $FOXP3^+$ T cell. In some embodiments, the T cell is a $CD4^+CD25^{lo}CD127^{hi}$ effector T cell. In some embodiments, the T cell is a $CD4^+CD25^{lo}CD127^{hi}CD45RA^{hi}CD45RO^-$ naïve T cell.

**[0096]** A T cell can be a genetically modified T cell. In some embodiments, the T cell receptor is a chimeric antigen receptor (CAR) containing a target-binding domain, a transmembrane domain, and an intracellular/endodomain/cytosolic domain. The cytosolic domain can contain one or more signaling domains and/or adaptor domains to provide robust T cell activation and anti-antigen activity. In some embodiments, the genetically modified T cell has mutated or heterologous) T cell receptor or a chimeric antigen receptor (CAR). For example, the T cell can have a CAR with one or more mutations to alter binding specificity or signaling. As yet another example, one or more of the T cell receptor domains can be replaced with a polypeptide having a different effector function.

**[0097]** As used herein, the term "pharmaceutically acceptable carrier" refers to an excipient or diluent in a pharmaceutical composition. The pharmaceutically acceptable carrier must be compatible with the other ingredients of the formulation and not deleterious to the recipient. In the present disclosure, the pharmaceutically acceptable carrier must provide adequate pharmaceutical stability to the active ingredient. The nature of the carrier differs with the mode of administration. For example, for intravenous administration, an aqueous solution carrier is generally used; for oral administration, a solid carrier is preferred.

**[0098]** The terms "patient," "subject," "individual," and the like are used interchangeably herein, and refer to any animal, *e.g.,* a mammal, such as a mouse, a primate, or a human. In certain non-limiting embodiments, the patient, subject or individual is a human.

**[0099]** The terms "treat" and "treatment" refer to both therapeutic treatment and prophylactic or preventive measures, wherein the object is to prevent or slow down an undesired physiological

change or disorder. For purpose of this disclosure, beneficial or desired clinical results include, but are not limited to, alleviation of symptoms, diminishment of extent of disease, stabilized (*i.e.,* not worsening) state of disease, delay or slowing of disease progression, amelioration or palliation of the disease state, and remission (whether partial or total), whether detectable or undetectable. "Treatment" can also mean prolonging survival as compared to expected survival if not receiving treatment. In other embodiments the terms "treat", "treatment" and "treating" refer to the inhibition of the progression of a proliferative disorder, either physically by, *e.g.,* stabilization of a discernible symptom, physiologically by, *e.g.,* stabilization of a physical parameter, or both. In other embodiments the terms "treat", "treatment" and "treating" refer to the reduction or stabilization of tumor size or cancerous cell count.

**[0100]** The term "effective amount" as used herein, refers to the amount of vectors of the present disclosure that is sufficient to produce genetically modified target cells that can effect treatment, prognosis, or diagnosis of a disease when the vectors are administered to a subject. An effective amount will vary depending upon the subject and disease condition being treated, the weight and age of the subject, the severity of the disease condition, the manner of administration and the like, which can readily be determined by one of ordinary skill in the art. Dosage regiments may be adjusted to provide the optimum therapeutic response. An effective amount is also one in which any toxic or detrimental effects (*i.e.,* side effects) caused by administration of the vectors are minimized and/or outweighed by the beneficial effects.

**[0101]** As used in herein, the singular forms "a", "an" and "the" include plural referents unless the content clearly dictates otherwise. Thus, for example, reference to "a CAR-T" optionally includes a combination of two or more such molecules, and the like.

**[0102]** The term "about" as used herein refers to the usual error range for the respective value readily known to the skilled person in this technical field. For example, for $K_D$ and $IC_{50}$ values ± 20%, ± 10%, or ± 5%, are within the intended meaning of the recited value.

**[0103]** Unless otherwise defined, all technical terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this disclosure belongs.

## DETAILED DESCRIPTION

### I. Introduction

**[0104]** The present disclosure provides methods and compositions related to the delivery of molecules into target cell nuclei for *in vivo* modification of the genome. While viruses and virally derived particles have the intrinsic capacity to deliver molecules to cells, the difficulty in engineering cell-type selectivity has hindered their use for therapeutic delivery. High treatment doses are typically required to improve transduction efficiency and are problematic because they can cause immune-mediated toxicity. The methods and compositions herein combine the use of different vectors that result in enhanced target cell recognition for delivery of genome-editing machinery. In some embodiments, due to enhanced specificity of the vectors and the short lifetime of protein gene-editing components (in contrast with prolonged expression of viral genomic material), the vectors enable the use of lower treatment doses, thereby reducing the toxic effects of off-target editing by minimizing the effective concentrations necessary for therapeutic benefit.

**[0105]** The targeting of transgenes to T cells using a polynucleotide-guided nuclease (which can include but is not limited to CRISPR-Cas nucleases) and donor template repair methods (e.g., homology-directed repair (HDR), homology-mediated end-joining (HMEJ), and homology-independent targeted integration (HITI)) has been useful for adoptive cell therapies and T cell biology. T cells have been engineered to express chimeric antigen receptors (CARs) for the treatment of treating hematological malignancies, and this approach may be extended to the treatment of solid tumors. As described in the examples, by using the combination of vector compositions and methods disclosed herein, targeted integration of a CAR at the TCRα constant (TRAC) locus under the control of the endogenous promoter conferred physiological receptor expression and yielded T cells with anti-tumor activity. Because the CRISPR-Cas nucleases and HDR gene-editing components are delivered by more than one vector to a target cell, *in vivo* gene modification takes place only when all the components are present. For example, a CRISPR-Cas nuclease and a guide polynucleotide (e.g., a gRNA) delivered by a membrane-derived vector, and an donor template polynucleotide delivered by an AAV vector, must be present in the same target cell nucleus to produce cleavage at a target gene and integration of the donor template polynucleotide into the target gene. Because both vectors must be present, and because each vector has a different mechanism for cell target selectivity, the methods herein for genome modification

are highly specific. While the methods and compositions herein are exemplified in T cells, they may be used for targeting any loci in different cell types, such as other immune cells, for the expression of any suitable heterologous polypeptide for the treatment of disease.

**[0106]** Thus in some embodiments, viral vectors (e.g., AAV vectors), optionally having altered tropism due to modifications in the capsid proteins, are provided in the methods and compositions, along with at least one second vector. These modified viral vectors not only have enhanced specificity for target cells, e.g., T cells, but can also have higher transduction rates, thereby enabling administration at lower doses. In some embodiments, the viral vectors are useful for delivering donor template polynucleotides containing a modification for insertion into the genome at a target locus. In some embodiments, the viral vectors may also deliver guide polynucleotides for use in combination with polynucleotide-guided nucleases. As discussed below, the polynucleotide-guided nucleases may be delivered by another vector. Thus, in many embodiments, the AAVs do not deliver the nuclease or a polynucleotide encoding the nuclease.

**[0107]** In some embodiments, the methods and compositions further comprise membrane-derived vectors (e.g., envelope delivery vehicles (EDVs) or lipid nanoparticles (LNPs)), which, in addition to a lipid membrane, comprise a viral envelope protein in the case of EDVs but lack self-replicating viral genetic material and thus are not infectious like a virus. Membrane-derived vectors can in some embodiments comprise externally displayed antibody fragments, conferring cell surface marker recognition of specific target cells. Antibody-antigen interactions are specific and these vectors can predictably deliver genome editing machinery to target cells of interest over bystander cells in mixed populations *in vivo*. In some embodiments, the membrane-derived vectors are useful for delivering polynucleotide-guided nucleases, and optionally, corresponding guide polynucleotides. These membrane-derived vectors are used in combination with the viral vectors discussed above, which deliver donor template polynucleotides. Thus, in many embodiments, the EDVs do not deliver the donor template polynucleotide.

**[0108]** Gene editing cannot occur if either the viral vector or the membrane-derived vector, and, consequently, if any of the three gene-editing components, is absent. The two different vectors are used in combination to bring together the required gene-editing components to the same target cell.

## II. Vectors

**[0109]** Disclosed herein are viral vectors (e.g., AAV vectors) useful for delivery of donor template polynucleotides to a target cell for the purposes of genetically modifying the genome of the target cell. Viral vectors are useful for delivery of large quantities of donor template polynucleotides to a cell that are sufficient for integration of a portion of a polynucleotide into genomic DNA via homology-directed repair. A viral vector disclosed herein comprises a capsid protein and donor template polynucleotides; the donor template polynucleotides are encapsidated by the viral vector. In some cases, the viral vector contains wild-type capsid proteins. In some cases, the capsid protein is modified to have enhanced specificity or altered tropism for a target cell type. In some embodiments, the viral vector also comprises a guide polynucleotide for delivery to a target cell with the donor template polynucleotide; the guide polynucleotide is encapsidated by the viral vector. In some embodiments, the donor template polynucleotide is a homology-directed repair template (HDRT) polynucleotide. In some embodiments, the donor template polynucleotide is a homology-independent targeted integration template (HITIT) polynucleotide. In some embodiments, the donor template polynucleotide is a homology-mediated end-joining template (HMEJT) polynucleotide. Viral vectors, capsid proteins, donor template polynucleotides, and guide polynucleotides are discussed in detail below.

**[0110]** Gene editing cannot take place using a viral vector alone and without the use of a membrane-derived vector comprising at least a polynucleotide-guided nuclease. A polynucleotide-guided nuclease is required for performing a cut or a nick on the target DNA to initiate the HDR process. Thus, also disclosed herein are membrane-derived vectors (e.g., EDVs and LNPs) useful for delivery of polynucleotide-guided nucleases to a target cell. In some embodiments, the membrane-derived vector is an enveloped delivery vehicle (EDV). An EDV disclosed herein comprises polynucleotide-guided nucleases; the polynucleotide-guided nucleases are encapsidated by the EDV. In some embodiments, the EDV also comprises guide polynucleotides; the guide polynucleotides are encapsidated by the EDV. In some embodiments where the polynucleotide-guided nuclease is a CRISPR-Cas nuclease, the nuclease forms a ribonucleoprotein complex (RNP) with a guide polynucleotide, e.g., a guide RNA (gRNA) and the RNP is encapsidated by the EDV. EDVs, polynucleotide-guided nucleases, and guide polynucleotides are discussed in detail below.

**[0111]** Alternatively, in some embodiments, the membrane-derived vector is a lipid nanoparticle (LNP). An LNP disclosed herein comprises polynucleotide-guided nucleases or nucleic acids encoding the polynucleotide-guided nuclease, e.g., the polynucleotide-guided nucleases, or nucleic acids encoding the nuclease, are encapsidated by the LNP. In some embodiments, the LNP also comprises guide polynucleotides, e.g., the guide polynucleotides are encapsidated by the LNP. In some embodiments where the polynucleotide-guided nuclease is a CRISPR-Cas nuclease, the nuclease forms an RNP with a guide polynucleotide, e.g., a gRNA; the RNP is encapsidated by the LNP. LNPs, polynucleotide-guided nucleases, and guide polynucleotides are discussed in detail below.

**[0112]** In some embodiments, in addition to the donor template polynucleotide, the AAV can also comprise a guide polynucleotide (e.g., gRNA) for directing a polynucleotide-guided nuclease (e.g., CRISPR-Cas nuclease) to a specific DNA locus. In some of these embodiments, the EDV/LNP comprises a polynucleotide-guided nuclease (e.g., CRISPR-Cas nuclease) and is without a guide polynucleotide (e.g., gRNA). Thus, in some embodiments, when the EDV/LNP and the AAV are present in the same target cell, the AAV guide polynucleotide (e.g., gRNA) can interact with the EDV/LNP polynucleotide-guided nuclease (e.g., CRISPR-Cas nuclease) and form an RNP.

**[0113]** In some embodiments, in addition to the donor template polynucleotide, the AAV can also comprise a guide polynucleotide (e.g., gRNA) for directing a polynucleotide-guided nuclease to a specific DNA locus. In some of these embodiments, the EDV/LNP comprises a polynucleotide-guided nuclease (e.g., CRISPR-Cas nuclease) and a guide polynucleotide (e.g., gRNA), and together, they form an RNP. In some embodiments, the EDV/LNP guide polynucleotide is a gRNA that directs the EDV/LNP polynucleotide-guided nuclease (e.g., CRISPR-Cas nuclease) to a specific DNA locus. In other embodiments, the EDV/LNP guide polynucleotide is a non-targeting gRNA – it does not direct the EDV/LNP polynucleotide-guided nuclease (e.g., CRISPR-Cas nuclease) to a specific DNA locus; rather, it functions to stabilize the RNP. Thus, in some embodiments, when the EDV/LNP and the AAV are present in the same target cell, the AAV guide polynucleotide (e.g., gRNA) can interact with the EDV/LNP polynucleotide-guided nuclease (e.g., CRISPR-Cas nuclease) and form an RNP. In some embodiments, the AAV

guide polynucleotide (e.g., gRNA) replaces the non-targeting EDV/LNP guide polynucleotide in the RNP.

**[0114]** In some embodiments, the AAV comprises a guide polynucleotide (e.g., gRNA) for directing a polynucleotide-guided nuclease (e.g., CRISPR-Cas nuclease) to a specific DNA locus, and the AAV is without a donor template polynucleotide. In some of these embodiments, the EDV/LNP comprises a polynucleotide-guided nuclease (e.g., CRISPR-Cas nuclease) and is without a guide polynucleotide (e.g., gRNA). Thus, in some embodiments, when the EDV/LNP and the AAV are present in the same target cell, the AAV guide polynucleotide (e.g., gRNA) can interact with the EDV/LNP polynucleotide-guided nuclease (e.g., CRISPR-Cas nuclease) and form an RNP. Because an AAV donor template polynucleotide is absent, DNA repair by HDR or HITI cannot take place. Thus, in these embodiments, the gene at the target locus is disrupted and and no heterologous sequence is inserted at the target locus.

#### a. Virus Vectors

**[0115]** In some embodiments, a virus vector (i.e., viral vector), used in conjunction with a membrane-derived vector (e.g., an EDV or an LNP) that delivers the nuclease, comprises a donor template polynucleotide for delivery to a target cell. In some embodiments, the virus vector further comprises a guide polynucleotide. In some embodiments, the virus vector is an adeno-associated virus (AAV) vector. In some embodiments, the virus vector, e.g., AAV vector, comprises a wild-type capsid protein or a capsid protein variant. In some embodiments, the virus vector comprises targeting polypeptides for selective binding to target cells.

##### i. Adeno-Associated Virus (AAV) Vectors

**[0116]** AAV, a member of the Parvovirus family, is a small, non-enveloped virus. Wild-type (WT) AAV is composed of an icosahedral protein capsid which encloses a single-stranded DNA genome. In WT AAVs, inverted terminal repeats (ITRs) flank the coding nucleotide sequences (e.g., a polynucleotides) for the non-structural proteins (encoded by Rep genes) and the structural proteins (encoded by capsid genes or Cap genes). Rep genes encode the non-structural proteins that regulate functions comprising the replication of the AAV genome. Cap genes encode the structural proteins, VP1, VP2 and/or VP3 that assemble to form the capsid. In some embodiments, the virus capsids can have one or more of any of the capsid proteins disclosed herein, including

WT and variant capsids. In some embodiments, the virus capsid can comprise AAV capsid proteins, or AAV capsid proteins in combination with any parvovirus capsid proteins.

**[0117]** In some embodiments, AAV vectors disclosed herein can be self-complementary AAV (scAAV) vectors. Self-complementary AAV (scAAV) vectors contain complementary sequences that are capable of spontaneously annealing (folding back on itself to form a double-stranded genome) when entering into infected cells, thus circumventing the need for converting a single-stranded DNA vector using the cell's DNA replication machinery. An AAV herein having a self-complementing genome can form a double stranded DNA molecule by virtue of its partially complementing sequences (e.g., complementing coding and non-coding strands of a transgene-encoding sequence).

**[0118]** In some embodiments, a wild-type or modified AAV vector (without modifications in tropism) may be used with a variety of target cells, e.g., T cells, B cells, NK cells, monocytes, macrophages, dendritic cells, or HSCs. In some embodiments, a modified AAV vector with enhanced or altered tropism for a particular target cell type (e.g., T cells, B cells, NK cells, monocytes, macrophages, dendritic cells, or HSCs) may be used. In some embodiments, the target cell is a T cell, and in some of those embodiments and the AAV vector is AAV6, AAV5, AAV9 or a variant thereof. In some of these embodiments, the vector is an AAV variant with an Ark315 capsid variant peptide sequence (SEQ ID NO: 9) or an Ark312 capsid variant peptide sequence (SEQ ID NO: 3), both of which are described below. In some embodiments, the target cell is a B cell, and in some of those embodiments the AAV vector is an AAV6 vector or a variant thereof. In some embodiments, the target cell is an NK cell, and in some of those embodiments the AAV vector is an AAV6 vector or a variant thereof. In some of these embodiments, the vector is an AAV variant with an Ark315 capsid variant peptide sequence (SEQ ID NO: 9) or an Ark312 capsid variant peptide sequence (SEQ ID NO: 3), both of which are described below. In some embodiments, the target cell is a monocyte or a macrophage, and in some of those embodiments the AAV vector is AAV6, AAV2, AAV9, or a variant thereof. In some embodiments, the target cell is a dendritic cell, and in some of those embodiments the AAV vector is AAV6, AAV2, AAV19, or a variant thereof. In some embodiments, the target cell is an HSC, and in some of those embodiments the AAV vector is AAV6 or AAV5 or a variant thereof.

**[0119]** In some embodiments, modified virus capsids herein can be a targeted virus capsid, comprising a targeting sequence (e.g., substituted or inserted in the viral capsid) that can direct the virus capsid to interact with cell-surface molecules present on desired target tissue(s). AAV capsid proteins, capsids and vectors comprising targeting sequences are described in, e.g., International Patent Application Publication No. WO 00/28004 and Hauck et al. (2003) J. Virology, 77:2768-2774); Shi et al. (2006) Human Gene Ther. 17:353-361 describing insertion of the integrin receptor binding motif RGD at positions 520 and/or 584 of the AAV capsid subunit; and U.S. Patent No. 7,314,912 describing insertion of the P1 peptide containing an RGD motif following amino acid positions 447, 534, 573 and 587 of the AAV2 capsid subunit). Other positions within the AAV capsid subunit that tolerate insertions are known in the art (e.g., positions 449 and 588 described by Grifman et al. (2001) Molecular Therapy 3:964-975).

**[0120]** In some embodiments, a virus capsid of the present disclosure can have relatively enhanced or reduced inefficient tropism toward certain cells of interest (e.g., immune cells). In some embodiments, a virus capsid has enhanced tropism toward target cells, e.g., T cells, NK cells, B cells, monocytes, macrophages, dendritic cells, or HSCs. A targeting sequence can advantageously be incorporated into AAV vectors of the present disclosure to thereby confer to the virus capsid a desired tropism and, optionally, selective tropism for particular tissue(s). AAV capsid proteins, capsids and vectors comprising targeting sequences are described, for example in International Patent Application Publication No. WO 2000/028004. As another example, one or more non-naturally occurring amino acids as described by Wang et al., *Annu Rev Biophys Biomol Struct.* 35:225-49 (2006) can be incorporated into an AAV capsid subunit of this disclosure at an orthogonal site as a means of redirecting a vector to desired target tissue(s). These unnatural amino acids can advantageously be used to chemically link molecules of interest to the AAV capsid protein including without limitation: glycans (mannose-dendritic cell targeting); RGD, bombesin or a neuropeptide for targeted delivery to specific cancer cell types; RNA aptamers or peptides selected from phage display targeted to specific cell surface receptors such as growth factor receptors, integrins, and the like. Methods of chemically modifying amino acids are known in the art (see, e.g., Greg T. Hermanson, *Bioconjugate Techniques,* 1st edition, Academic Press, 1996). In some embodiments, the targeting sequence can be a virus capsid sequence (e.g., an autonomous parvovirus capsid sequence, AAV capsid sequence, or any other viral capsid sequence) that directs infection to a particular cell type(s).

**[0121]** In some embodiments, capsid protein variants, virus capsids, and/or virus vectors disclosed herein can have equivalent or enhanced tropism relative to the tropism of the AAV serotype from which capsid protein variant, virus capsid and/or vector originated. In some embodiments, capsid protein variants, virus capsids, and/or virus vectors disclosed herein can have an altered or different tropism relative to the tropism of the AAV serotype from which the capsid protein variant, virus capsid and/or vector originated. In some embodiments, capsid protein variants, virus capsids, and/or virus vectors disclosed herein can have or be engineered to have tropism or enhanced tropism for immune cells (e.g., T cells, NK cells, B cells, monocytes, macrophages, or dendritic cells) or HSCs.

**[0122]** In some embodiments, capsid protein variants, virus capsids, and/or AAV vectors disclosed herein can produce an attenuated immunological response relative to the immunological response of the AAV serotype from which the capsid protein variant, virus capsid and/or vector originated. In some embodiments, capsid protein variants, virus capsids, and/or AAV vectors disclosed herein can be administered to a subject in multiple dosages (e.g., about two doses, about three doses, about four doses, about 5 doses, about 10 doses, about 15 doses, about 20 doses, about 40 doses, as many doses as needed to observe one or more desired responses) relative to the number of doses that can be administered using the AAV serotype from which the capsid protein variant, virus capsid and/or vector originated.

**[0123]** In some embodiments, capsid protein variants, virus capsids, and/or AAV vectors disclosed herein can be modified to display targeting polypeptides on its surface. The targeting polypeptides can bind to a target that is expressed on the surface of a target cell (e.g., T cell, NK cell, B cell, monocyte, macrophage, dendritic cell, or HSC), including ligands and cell receptors. In some embodiments, the targeting polypeptide is covalently linked to a capsid protein, e.g., VP1, VP2, and/or VP3. In some embodiments, the targeting polypeptide comprises an antigen-binding domain or a fragment thereof. Non-limiting examples of antigen-binding domains include an scFv, a diabody, a triabody, a nanobody, any antigen-binding fragment comprising the $V_H$ and $V_L$ domains of an antibody, a bispecific antibody, an affibody, an affilin, an affimer, an affitin, an αbody, an anticalin, an avimer, a designed ankyrin repeat protein (DARPin), a Fynomer, a Kunitz domain peptide, a monobody, a repebody, a VLR, and a nanoCLAMP. In some embodiments, targeting polypeptide comprises a VHH. In some embodiments, the targeting polypeptide

comprises a DARPin. AAVs with targeting polypeptides, e.g., antibodies or DARPins, are discussed in Michels, Alexander et al. "Lentiviral and adeno-associated vectors efficiently transduce mouse T lymphocytes when targeted to murine CD8." Molecular therapy. Methods & clinical development vol. 23 334-347. 1 Oct. 2021, doi:10.1016/j.omtm.2021.09.014; and Eichhoff, Anna Marei et al. "Nanobody-Enhanced Targeting of AAV Gene Therapy Vectors." Molecular therapy. Methods & clinical development vol. 15 211-220. 16 Sep. 2019, doi:10.1016/j.omtm.2019.09.003.

### 1. Capsid Modifications

**[0124]** In some embodiments, the virus vector comprises AAV capsid proteins (VP1, VP2 and/or VP3) comprising a modification (e.g., a substitution and/or deletions) in the amino acid sequence relative to a WT capsid protein, and AAV capsids and AAV vectors comprising the modified AAV capsid protein. As detailed herein, modifications of disclosed capsid proteins can confer one or more desirable properties to virus vectors comprising the modified AAV capsid protein variants herein, including without limitation, the ability to evade neutralizing antibodies and/or the ability to specifically and selectively target a cell or tissue of interest, as compared to wild-type capsid. Examples of AAV vectors, AAV capsid proteins, and suitable mutations are described in, e.g., International Patent Application Publication Nos. WO 2023/004407 and WO 2022/155482.

**[0125]** In some embodiments, AAV vectors herein can be engineered to include one or more capsid protein variants. In some embodiments, AAV vectors herein can be engineered to include at least one or more amino acid substitutions, wherein the one or more substitutions can modify one or more antigenic sites on the AAV capsid protein. The modification of the one or more antigenic sites can result in reduction or inhibition of binding by an antibody to the one or more antigenic sites and/or inhibition of neutralization of infectivity of a virus particle comprising said a capsid protein variant herein.

**[0126]** In some embodiments, the AAV capsid protein variant comprises one or more amino acid modifications (e.g., substitutions and/or deletions), wherein the one or more modifications modify one or more antigenic sites on the AAV capsid protein. In some embodiments, modification of the one or more antigenic sites can result in inhibition or reduction of binding by an antibody to the one or more antigenic sites and/or inhibition of neutralization of infectivity of a

virus particle comprising said AAV capsid protein. In some embodiments, the modified antigenic site can prevent or reduce antibodies from binding or recognizing or neutralizing AAV capsids. In some embodiments, the antibody can be an IgG (including IgG1, IgG2a, IgG2b, IgG3), IgM, IgE or IgA. In some embodiments, the modified antigenic site can prevent or reduce binding, recognition, or neutralization of AAV capsids by antibodies from different animal species, wherein the animal is human, canine, porcine, bovine, non-human primate, rodent (e.g., mouse), feline, or equine.

**[0127]** In some embodiments, modification of the one or more antigenic sites can result in tropism of the AAV vectors herein to one or more cell types, one or more tissue types, or any combination thereof. As used herein, "tropism" refers to preferential entry of the virus into certain cells or tissues, optionally followed by expression (e.g., transcription and, optionally, translation) of a sequence(s) carried by the viral genome in the cell, e.g., for a recombinant virus, expression of a heterologous nucleic acid(s) of interest. In some embodiments, modification of the one or more antigenic sites can result in AAV vectors herein that can exhibit tropism for one or more cell types and/or tissues throughout the body of a subject. In some embodiments, modification of the one or more antigenic sites can result in AAV vectors herein that can exhibit tropism to one or more immune cell types. In some embodiments, modification of the one or more antigenic sites can result in AAV vectors herein that can exhibit tropism to T-cells (CD4 T cells and/or CD8 T cells), monocytes, macrophages, dendritic cells, HSCs, B cells, and/or natural killer (NK) cells. In some embodiments, modification of the one or more antigenic sites can result in AAV vectors herein that can exhibit tropism to T cells and NK cells.

**[0128]** In some embodiments, AAV capsid protein variants disclosed herein can have an amino acid sequence with about 85% (e.g., about 85%, 90%, 95%, 99%, 100%) similarity to a naturally occurring capsid protein. In some embodiments, a naturally occurring capsid protein herein can be derived from a single species. In some embodiments, AAV capsid protein variants having at least one amino acid substitution as disclosed herein can have an amino acid sequence with about 85% (e.g., about 85%, 90%, 95%, 99%, 100%) similarity to a naturally occurring capsid protein having an amino acid sequence referenced by GenBank Accession Numbers: NC_002077, NC_001401, NC_001729, NC_001863, NC_001829, NC_001862, NC_000883, NC_001701, NC_001510, NC_006152, NC_006261, AF063497, U89790, AF043303, AF028705, AF028704, J02275,

J01901, J02275, X01457, AF288061, AH009962, AY028226, AY028223, NC_001358, NC_001540, AF513851, AF513852, AY530579, and any combination thereof.

**[0129]** In some embodiments, AAV capsid protein variants disclosed herein can have at least one amino acid substitution that can replace any seven amino acids in an AAV capsid protein from a serotype having tropism for immune cells (e.g., T cells, NK cells, B cells, monocytes, macrophages, or dendritic cells) or HSCs. In some embodiments, AAV capsid protein variants disclosed herein can have at least one amino acid substitution that can replace any seven amino acids in an AAV capsid protein of any AAV serotype having tropism for T cells. In some embodiments, AAV capsid protein variants disclosed herein can have at least one amino acid substitution that can replace any seven amino acids in an AAV capsid protein of any AAV serotype having tropism for NK cells.

**[0130]** In some embodiments, AAV capsid protein variants herein or fragments thereof can have an amino acid sequence with about 85% (e.g., about 85%, 90%, 95%, 99%, 100%) similarity to a naturally occurring VP1 capsid protein or fragment thereof. In some embodiments, capsid protein variants herein can comprise an amino acid substitution at one or more (e.g., 2, 3, 4, 5, 6, or 7) of amino acid residues 454-460 of AAV6 (VP1 numbering), in any combination, or the equivalent amino acid residues in AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV10, AAV11, AAV12, AAV19, AAVrh8, AAVrh10, AAVrh32.33, bovine AAV or avian AAV.

**[0131]** In some embodiments, capsid protein variants herein can have at least 90% (e.g., about 90%, 95%, 99%, 100%) sequence identity to the native sequence of the AAV6 capsid (WT AAV6; SEQ ID NO: 1; shown below). In some embodiments, capsid protein variants herein can comprise a substitution at one or more (e.g., 2, 3, 4, 5, 6, or 7) amino acid residues within a SEQ ID NO: 2 (454-GSAQNKD-460 (VP1 numbering)) on the capsid surface of AAV6 (SEQ ID NO: 1) in any combination. The amino acid corresponding to position 454G of SEQ ID NO: 1 can be any amino acid other than G. The amino acid corresponding to position 455S of SEQ ID NO: 1 can be any amino acid other than S. The amino acid corresponding to position 456A of SEQ ID NO: 1 can be any amino acid other than A. The amino acid corresponding to position 457Q of SEQ ID NO: 1 can be any amino acid other than Q. The amino acid corresponding to position 458N of SEQ ID NO: 1 can be any amino acid other than N. The amino acid corresponding to position 459K of

SEQ ID NO: 1 can be any amino acid other than K. The amino acid corresponding to position 460D of SEQ ID NO: 1 can be any amino acid other than D. Non-limiting examples of AAV capsid proteins, including AAV6 capsid, are discussed in International Application Publication No. WO 2023/004407.

**[0132]** In some embodiments, AAV vectors herein comprise (i) an AAV6 capsid protein variant and (ii) a donor template polynucleotide. In some embodiments, AAV vectors herein comprise (i) an AAV6 capsid protein variant and (ii) a donor template polynucleotide, wherein the capsid protein variant can comprise a peptide having any one of the sequences in Table 1 below at amino acids 454-460 (VP1 numbering) of a WT AAV6 capsid protein, (SEQ ID NO: 1).

**[0133]** In some embodiments, capsid protein variants herein can comprise a peptide wherein the amino acids corresponding to amino acid position 454-460 (VP1 numbering) of a WT AAV6 capsid protein (SEQ ID NO: 2), can be substituted with amino acids corresponding to any one of SEQ ID NO: 3–57 and 224-263. Table 1 below provides amino acids corresponding to any one of SEQ ID NO: 3–57 and 224-263 (or AAV6 WT peptide sequence shown as SEQ ID NO: 2).

| **Table 1– Listing of AAV6 Capsid Variants (with SEQ ID NOs)** | | |
|---|---|---|
| **Designation** | **Sequence (positions 454-460 relative to SEQ ID NO: 1)** | **SEQ ID NO:** |
| **AAV6 (WT)** | **GSAQNKD (G454 S455 A456 Q457 N458 K459 D460S)** | **2** |
| Ark312 | HAPRVEE | 3 |
| Ark314 | SAPRSEN | 4 |
| Ark319 | MAPRQEG | 5 |
| Ark320 | AAPRSET | 6 |
| Ark322 | QAPRQEG | 7 |
| Ark327 | VAPRDPA | 8 |
| Ark315 | VAPRDSS | 9 |
| Ark316 | EAPRAMS | 10 |
| Ark317 | LAPRDRE | 11 |
| Ark318 | EAPRSVS | 12 |
| Ark321 | EAPRESA | 13 |

| | | |
|---|---|---|
| Ark323 | DAPRLGG | 14 |
| Ark324 | PAPRESS | 15 |
| Ark326 | LAPREAL | 16 |
| Ark328 | LAPREWE | 17 |
| Ark329 | VAPRDAV | 18 |
| Ark330 | DAPRQTM | 19 |
| Ark331 | EAPRQTS | 20 |
| Ark332 | EAPRLDA | 21 |
| Ark333 | MAPRVVD | 22 |
| Ark334 | TAPRETN | 23 |
| Ark335 | MAPRDAG | 24 |
| Ark336 | EAPRAVQ | 25 |
| Ark337 | VAPREVM | 26 |
| Ark338 | EAPRQPT | 27 |
| Ark339 | NAPREVN | 28 |
| Ark340 | PAPRDGM | 29 |
| Ark341 | EAPRDVI | 30 |
| Ark342 | EAPRFGG | 31 |
| Ark343 | EAPRSQW | 32 |
| Ark344 | IAPRGSD | 33 |
| Ark345 | PAPRMIE | 34 |
| Ark346 | IAPRSVG | 35 |
| Ark347 | FPAGSME | 36 |
| Ark348 | EAPRSRD | 37 |
| Ark349 | EAPRRLE | 38 |
| Ark350 | VAPRVAG | 39 |
| Ark351 | GLGLGNG | 40 |
| Ark352 | YVHSAEE | 41 |
| Ark353 | DAPRIVK | 42 |

| | | |
|---|---|---|
| Ark354 | EAPRVWS | 43 |
| Ark355 | QGPGLLG | 44 |
| Ark356 | SAPRDHG | 45 |
| Ark357 | YAPRELG | 46 |
| Ark358 | LAPREPG | 47 |
| Ark359 | VAPRAGA | 48 |
| Ark360 | DAPRVQL | 49 |
| Ark361 | AAPRSVA | 50 |
| Ark362 | SAPREME | 51 |
| Ark363 | QTDAQPR | 52 |
| Ark325 | -KEGA- - (dashes are not present, meaning positions 454, 459 and 460 are absent) | 53 |
| Ark396 | NAPREVN | 54 |
| Ark397 | PVGDVGK | 55 |
| Ark600 | VQKGSYE | 56 |
| Ark601 | RPDGNSL | 57 |
| Ark602 | KEGQAIQ | 224 |
| Ark603 | KGSEGIS | 225 |
| Ark604 | AKEVERH | 226 |
| Ark605 | SEKGASS | 227 |
| Ark606 | SSSGDNM | 228 |
| Ark607 | KSTNDLS | 229 |
| Ark608 | LDSKRLE | 230 |
| Ark609 | EAPRESA | 231 |
| Ark610 | SETEARK | 232 |
| Ark611 | GVADGVK | 233 |
| Ark612 | GPDDTRK | 234 |
| Ark613 | SEKMGFN | 235 |
| Ark614 | RSTESIE | 236 |

| | | |
|---|---|---|
| Ark615 | LQKDGAA | 237 |
| Ark616 | SSTVERV | 238 |
| Ark617 | LAPRSPD | 239 |
| Ark618 | GSERPNV | 240 |
| Ark619 | AAPRECG | 241 |
| Ark620 | STDRQRD | 242 |
| Ark621 | ASTGEGR | 243 |
| Ark622 | VASGSPS | 244 |
| Ark623 | VSKDVPN | 245 |
| Ark624 | SNEVKIG | 246 |
| Ark625 | TTGTDRG | 247 |
| Ark626 | VGSDGVR | 248 |
| Ark627 | LEGKERG | 249 |
| Ark628 | TTTADKA | 250 |
| Ark629 | SGVREGV | 251 |
| Ark630 | ADRDASA | 252 |
| Ark631 | TNTSDKL | 253 |
| Ark632 | GDGKGQW | 254 |
| Ark633 | AVDKERL | 255 |
| Ark634 | PTAKDMA | 256 |
| Ark635 | HQEGERM | 257 |
| Ark636 | REDDGRC | 258 |
| Ark637 | KEEASRH | 259 |
| Ark638 | SVSVDRH | 260 |
| Ark639 | TDKDGMS | 261 |
| Ark640 | TESGALS | 262 |
| Ark641 | HDIDSPR | 263 |

**[0134]** In some embodiments, capsid protein variants herein can share at least about 85% (e.g., about 85%, 90%, 95%, 99%, or 100%) amino acid sequence similarity with any one of the sequences set forth in SEQ ID NO: 1 (WT AAV6) or SEQ ID NO: 58 (Ark312), provided below. Underlined and bold amino acids below indicate position 454-460 that may be substituted with any of the sequences in the tables above.

WT AAV6 (SEQ ID NO: 1) –

MAADGYLPDWLEDNLSEGIREWWDLKPGAPKPKANQQKQDDGRGLVLPGYKYLGPF
NGLDKGEPVNAADAAALEHDKAYDQQLKAGDNPYLRYNHADAEFQERLQEDTSFGGN
LGRAVFQAKKRVLEPFGLVEEGAKTAPGKKRPVEQSPQEPDSSSGIGKTGQQPAKKRLN
FGQTGDSESVPDPQPLGEPPATPAAVGPTTMASGGGAPMADNNEGADGVGNASGNWH
CDSTWLGDRVITTSTRTWALPTYNNHLYKQISSASTGASNDNHYFGYSTPWGYFDFNRF
HCHFSPRDWQRLINNNWGFRPKRLNFKLFNIQVKEVTTNDGVTTIANNLTSTVQVFSDS
EYQLPYVLGSAHQGCLPPFPADVFMIPQYGYLTLNNGSQAVGRSSFYCLEYFPSQMLRT
GNNFTFSYTFEDVPFHSSYAHSQSLDRLMNPLIDQYLYYLNRTQNQS**GSAQNKD**LLFSR
GSPAGMSVQPKNWLPGPCYRQQRVSKTKTDNNNSNFTWTGASKYNLNGRESIINPGTA
MASHKDDKDKFFPMSGVMIFGKESAGASNTALDNVMITDEEEIKATNPVATERFGTVA
VNLQSSSTDPATGDVHVMGALPGMVWQDRDVYLQGPIWAKIPHTDGHFHPSPLMGGF
GLKHPPPQILIKNTPVPANPPAEFSATKFASFITQYSTGQVSVEIEWELQKENSKRWNPEV
QYTSNYAKSANVDFTVDNNGLYTEPRPIGTRYLTRPL

Ark312 (SEQ ID NO: 58) –

MAADGYLPDWLEDNLSEGIREWWDLKPGAPKPKANQQKQDDGRGLVLPGYKYLGPF
NGLDKGEPVNAADAAALEHDKAYDQQLKAGDNPYLRYNHADAEFQERLQEDTSFGGN
LGRAVFQAKKRVLEPFGLVEEGAKTAPGKKRPVEQSPQEPDSSSGIGKTGQQPAKKRLN
FGQTGDSESVPDPQPLGEPPATPAAVGPTTMASGGGAPMADNNEGADGVGNASGNWH
CDSTWLGDRVITTSTRTWALPTYNNHLYKQISSASTGASNDNHYFGYSTPWGYFDFNRF
HCHFSPRDWQRLINNNWGFRPKRLNFKLFNIQVKEVTTNDGVTTIANNLTSTVQVFSDS
EYQLPYVLGSAHQGCLPPFPADVFMIPQYGYLTLNNGSQAVGRSSFYCLEYFPSQMLRT
GNNFTFSYTFEDVPFHSSYAHSQSLDRLMNPLIDQYLYYLNRTQNQS**HAPRVEE**LLFSR
GSPAGMSVQPKNWLPGPCYRQQRVSKTKTDNNNSNFTWTGASKYNLNGRESIINPGTA
MASHKDDKDKFFPMSGVMIFGKESAGASNTALDNVMITDEEEIKATNPVATERFGTVA
VNLQSSSTDPATGDVHVMGALPGMVWQDRDVYLQGPIWAKIPHTDGHFHPSPLMGGF
GLKHPPPQILIKNTPVPANPPAEFSATKFASFITQYSTGQVSVEIEWELQKENSKRWNPEV
QYTSNYAKSANVDFTVDNNGLYTEPRPIGTRYLTRPL

**[0135]** In some embodiments, the AAV capsid comprises a variant, e.g., as described above, resulting in enhanced tropism, e.g., for a human immune cell (e.g., a T-cell or NK cell). In some embodiments, the AAV capsid protein variant further or alternatively comprises one or more exogenous targeting sequence that alters the tropism of a virus capsid or virus vector comprising the modified AAV capsid protein. In some embodiments, the targeting sequence is derived from a

naturally-occurring targeting peptide or protein. In some embodiments, targeting sequences can include ligands and other peptides that bind to cell surface receptors and glycoproteins, such as arginine-glycine-aspartate (RGD) peptide sequences, bradykinin, hormones, peptide growth factors (e.g., epidermal growth factor, nerve growth factor, fibroblast growth factor, platelet-derived growth factor, insulin-like growth factors I and II, etc.), cytokines, melanocyte stimulating hormone (e.g., α, β or γ), neuropeptides and endorphins, and the like, and fragments thereof that retain the ability to target cells to their cognate receptors. Other illustrative peptides and proteins include, but are not limited to substance P, keratinocyte growth factor, neuropeptide Y, gastrin releasing peptide, interleukin 2, hen egg white lysozyme, erythropoietin, gonadoliberin, corticostatin, β-endorphin, leu-enkephalin, rimorphin, α-neo-enkephalin, angiotensin, pneumadin, vasoactive intestinal peptide, neurotensin, motilin, and fragments thereof as described above. As yet a further alternative, the binding domain from a toxin (e.g., tetanus toxin or snake toxins, such as α-bungarotoxin, and the like) can be substituted into the capsid protein as a targeting sequence. In some embodiments, a AAV capsid protein herein can be modified by substitution of a "nonclassical" import/export signal peptide (e.g., fibroblast growth factor-1 and -2, interleukin 1, HIV-1 Tat protein, herpes virus VP22 protein, and the like) as described by Cleves (*Current Biology* 7:R318 (1997)) into the AAV capsid protein. In some embodiments, a targeting sequence for use herein can be a peptide that can be used for chemical coupling (e.g., can comprise arginine and/or lysine residues that can be chemically coupled through their R groups) to another molecule that targets entry into a cell.

### 2. AAV Vector Components and Methods of Producing AAV Vectors

**[0136]** In some embodiments, AAV vectors disclosed herein can comprise further elements useful for expression, such as at least one suitable promoter which controls the expression of the transgene-encoding sequence. Such a promoters can be ubiquitous, tissue-specific, strong, weak, regulated, chimeric, etc., to allow efficient and suitable production of the protein in the infected tissue. The promoter can be homologous to the encoded protein, or heterologous, including cellular, viral, fungal, plant or synthetic promoters. Most preferred promoters for use herein can be functional in human cells. Non-limiting examples of ubiquitous promoters include viral promoters, particularly the CMV promoter, the RSV promoter, the SV40 promoter, etc. and

cellular promoters such as the PGK (phosphoglycerate kinase) promoter. In some embodiments, viral promoters herein can be a CMV promoter, a SV40 promoter, or any combination thereof.

**[0137]** In some embodiments, AAV vectors disclosed herein can comprise further elements necessary for expression, such as at least one suitable promoter which controls the expression of the transgene-encoding sequence after infection of the appropriate cells. Suitable promoters for use herein include, in addition to the AAV promoters, e.g. the cytomegalovirus (CMV) promoter or the chicken beta actin/cytomegalovirus hybrid promoter (CAG), an endothelial cell-specific promoter such as the VE-cadherin promoter, as well as steroid promoters and metallothionein promoters. In some embodiments, the promoter used in the vectors disclosed herein can be a CAG promoter.

**[0138]** In some embodiments, a disclosed transgene-encoding sequence can comprise a tissue specific promoter which is functionally linked to the transgene-encoding sequence to be expressed. Accordingly, the specificity of the vectors according to the disclosure for the tissue (e.g., immune cells such as T cells, NK cells, B cells, monocytes, macrophages, or dendritic cells; or HSCs) can be further increased. In some embodiments, a vector disclosed herein can have a tissue-specific promoter whose activity in the specific tissue is at least about 2-fold, 5-fold, 10-fold, 20-fold, 50-fold or 100-fold higher than in a tissue which is not the specific tissue. In some embodiments, a tissue specific promoter herein is a human a tissue specific promoter. In some embodiments, the expression cassette can also include an enhancer element for increasing the expression levels of exogenous protein to be expressed. Furthermore, the expression cassette can further comprise polyadenylation sequences, such as the SV40 polyadenylation sequences or polyadenylation sequences of bovine growth hormone.

**[0139]** In some embodiments, an AAV vector disclosed herein can include a modified capsid, including proteins or peptides of non-viral origin or structurally modified, to alter the tropism of the vector. For example, the capsid can include a ligand of a particular receptor, or a receptor of a particular ligand, to target the vector towards cell type(s) expressing said receptor or ligand, respectively.

**[0140]** In some embodiments, AAV vectors disclosed herein can be prepared or derived from various serotypes of AAVs. The term "serotype" is a distinction with respect to an AAV having a capsid which is serologically distinct from other AAV serotypes. Serologic distinctiveness is

determined on the basis of the lack of cross-reactivity between antibodies to the AAV as compared to other AAV. Cross-reactivity can be measured using methods known in the art. For example, cross-reactivity herein can be measured using a neutralizing antibody assay. For this assay polyclonal serum is generated against a specific AAV in a rabbit or other suitable animal model using the adeno-associated viruses. In this assay, the serum generated against a specific AAV is then tested in its ability to neutralize either the same (homologous) or a heterologous AAV. The dilution that achieves 50% neutralization is considered the neutralizing antibody titer. If for two AAVs the quotient of the heterologous titer divided by the homologous titer is lower than 16 in a reciprocal manner, those two vectors are considered as the same serotype. Conversely, if the ratio of the heterologous titer over the homologous titer is 16 or more in a reciprocal manner the two AAVs are considered distinct serotypes.

**[0141]** In some embodiments, AAV vectors herein can be mixed of at least two serotypes of AAVs or with other types of viruses to produce chimeric (e.g., pseudotyped) AAV viruses. In some embodiments, AAV vectors herein can be a human serotype AAV vector. Such a human AAV can be derived from any known serotype, e.g., from any one of serotypes 1-11.

**[0142]** In some embodiments, AAV vector genomes described herein can be packaged into virus particles which can be used to deliver the genome for transgene-encoding sequence expression in target cells. In some embodiments, AAV vector genomes disclosed herein can be packaged into particles by transient transfection, use of producer cell lines, combining viral features into Ad-AAV hybrids, use of herpesvirus systems, or production in insect cells using baculoviruses.

**[0143]** A method of generating a packaging cell for use herein can involve creating a cell line that stably expresses all the necessary components for AAV particle production. For example, a plasmid (or multiple plasmids) comprising a rAAV genome lacking AAV rep and cap genes, AAV rep and cap genes separate from the rAAV genome, and a selectable marker, such as a neomycin resistance gene, are integrated into the genome of a cell. AAV genomes have been introduced into bacterial plasmids by procedures such as GC tailing, addition of synthetic linkers containing restriction endonuclease cleavage sites, or by direct, blunt-end ligation. The packaging cell line is then infected with a helper virus, such as adenovirus. The advantages of this method are that the cells are selectable and are suitable for large-scale production of rAAV. Examples of suitable

methods herein employ adenovirus or baculovirus, rather than plasmids, to introduce rAAV genomes and/or rep and cap genes into packaging cells.

**[0144]** In some embodiments, AAV vectors and/or AAV particles herein can have one or more improvements compared to naturally isolated AAV vectors. As used herein, a "naturally isolated AAV vector" refers to a vector that does not comprise one or more of the capsid protein variants disclosed herein. In some embodiments, AAV vectors and/or AAV particles herein can have increased gene transfer efficiency in a cell compared to naturally isolated AAV vectors. In some embodiments, AAV vectors and/or AAV particles herein can have at least about 2-fold to about 50-fold (e.g., about 2-, 4-, 6-, 8-, 10-, 20-, 30-, 40-, 50-fold) increased gene transfer efficiency in a cell compared to naturally isolated AAV vectors.

**[0145]** In some embodiments, AAV vectors and/or AAV particles herein can have increased gene transfer efficiency in the cell and/or tissue of one or more mammalian species. In some embodiments, AAV vectors and/or AAV particles herein can have increased gene transfer efficiency in the cell and/or tissue of a human, e.g., a human T-cell.

**[0146]** In some embodiments, AAV vectors and/or AAV particles herein can have a higher vector titer compared to naturally isolated AAV vectors. In some embodiments, AAV vectors and/or AAV particles herein can have at least about 2-fold to about 50-fold (e.g., about 2-, 4-, 6-, 8-, 10-, 20-, 30-, 40-, 50-fold) higher vector titer compared to naturally isolated AAV vectors.

**[0147]** In some embodiments, AAV vectors and/or AAV particles herein can be less susceptible to antibody-mediated neutralization compared to naturally isolated AAV vectors. In some embodiments, AAV vectors and/or AAV particles herein can be less susceptible to antibody-mediated neutralization by about 2-fold to about 50-fold (e.g., about 2-, 4-, 6-, 8-, 10-, 20-, 30-, 40-, 50-fold) compared to naturally isolated AAV vectors. In some embodiments, AAV vectors and/or AAV particles herein can be less susceptible to antibody-mediated neutralization for at least about 1 hour to about 24 hours (e.g., about 1, 2, 4, 8, 12, 16, 20, 24 hours) after administration to a subject compared to naturally isolated AAV vectors.

**[0148]** In some embodiments, AAV vectors and/or AAV particles herein can produce lower levels of anti-AAV antibodies after at least one administration to a subject herein compared to naturally isolated AAV vectors. In some embodiments, AAV and/or AAV particles herein can

produce about 2-fold to about 50-fold (e.g., about 2-, 4-, 6-, 8-, 10-, 20-, 30-, 40-, 50-fold) less anti-AAV antibodies after at least one administration to a subject herein compared to naturally isolated AAV vectors In some embodiments, gene therapies comprising AAV vectors and/or AAV particles herein can be administered about 2 times to about 10 times (e.g., about 2, 3, 4, 5, 6, ,7, 8, 9, 10) to a subject herein without becoming susceptible to antibody-mediated neutralization.

### ii. Exemplary AAVs

**[0149]** In some embodiments, the AAV vector comprises a capsid variant with the peptide sequence HAPRVEE (SEQ ID NO: 3) at positions corresponding to amino acids 454-460 (VP1 numbering) of a native AAV6 capsid protein, (SEQ ID NO: 1).

**[0150]** In some embodiments, the AAV vector comprises an HDRT polynucleotide (e.g., comprising or composed of DNA). HDRT polynucleotides are discussed in detail below. In some embodiments, the HDRT polynucleotide comprises a first TRAC locus homology sequence (SEQ ID NO: 205). In some embodiments, the HDRT polynucleotide comprises a second TRAC locus homology sequence (SEQ ID NO: 206). In some embodiments, the HDRT polynucleotide comprises 1, 2, or 3 coding sequences for 1, 2, or 3 P2A peptide sequences chosen from SEQ ID NOS: 207-209. In some embodiments, the HDRT polynucleotide comprises a coding sequence for a CAR designated as "19-166-28z 1XX" or "1928z-1XX" (SEQ ID NO: 210). In some embodiments, the HDRT polynucleotide comprises a coding sequence for an EGFRT polypeptide sequence (SEQ ID NO: 211). In some embodiments, the HDRT polynucleotide comprises the sequence as set forth in SEQ ID NO: 212. The HDRT polynucleotide-related sequences herein are described below in detail.

**[0151]** In some embodiments, the AAV vector also comprises a guide polynucleotide, e.g., a guide RNA (gRNA). In some embodiments, the gRNA is targeted to the *TRAC* locus. In some embodiments, the gRNA is targeted to exon 1 of the *TRAC* locus. In some embodiments, the gRNA comprises the sequence of CAGGGTTCTGGATATCTGT (SEQ ID NO: 137) or TCAGGGTTCTGGATATCTGT (SEQ ID NO: 138). In some embodiments, the gRNA is under control of the U6 promoter. In some embodiments, the AAV vector comprises a U6 promoter-TRAC-sgRNA scaffold comprises the sequence as set forth in SEQ ID NO: 139.

### b. Envelope Delivery Vehicles (EDVs)

**[0152]** Disclosed herein are envelope delivery vehicles (EDVs) for use in conjunction with a virus vector (e.g., an AAV vector) of the present disclosure. An EDV is a membrane-derived vector that comprises at least a group-specific antigen (gag) protein fragment and a viral envelope protein. An EDV does not contain viral genetic material and cannot self-propagate and is not infectious. In some embodiments, modified lentiviruses are used to package EDV cargo molecules and the lentivirus are then used to transduce host cells for production of the EDVs with encapsidated cargo molecules. Examples of EDVs and methods for producing EDVs are described in Hamilton, Jennifer R et al. "Targeted delivery of CRISPR-Cas9 and transgenes enables complex immune cell engineering." Cell reports vol. 35,9 (2021): 109207. doi:10.1016/j.celrep.2021.109207; Hamilton, Jennifer R et al. "In vivo human T cell engineering with enveloped delivery vehicles." Nature biotechnology, 10.1038/s41587-023-02085-z. 11 Jan. 2024, doi:10.1038/s41587-023-02085-z; and U.S. Patent Application Publication No. 2022/0403379.

**[0153]** In some embodiments, the EDV comprises a polynucleotide-guided nuclease for delivery to a target cell. Many suitable gene-editing nucleases may be used as a polynucleotide-guided nuclease and are discussed in detail below. Suitable nucleases include, but are not limited to, a homing nuclease polypeptide; a FokI polypeptide; a transcription activator-like effector nuclease (TALEN) polypeptide; a MegaTAL polypeptide; a meganuclease polypeptide; a zinc finger nuclease (ZFN); an ARCUS nuclease; and the like. The meganuclease can be engineered from an LADLIDADG homing endonuclease (LHE). A megaTAL polypeptide can comprise a TALE DNA binding domain and an engineered meganuclease. See, e.g., WO 2004/067736 (homing endonuclease); Urnov et al. (2005) Nature 435:646 (ZFN); Mussolino et al. (2011) Nucle. Acids Res. 39:9283 (TALE nuclease); Boissel et al. (2013) Nucl. Acids Res. 42:2591 (MegaTAL). In some embodiments, the polynucleotide-guided nuclease is a CRISPR-Cas nuclease. In some embodiments, the polynucleotide-guided nuclease is CRISPR-Cas9. In some embodiments, the polynucleotide-guided nuclease is a CRISPR-associated transposase (CAST). In some embodiments, the polynucleotide-guided nuclease, optionally comprising nuclease-inactivating mutations (e.g., dCas9), is covalently-linked with a reverse transcriptase. In some embodiments, the polynucleotide-guided nuclease is a CRISPR-directed integrase (e.g., CRISPR-Cas9 nickase covalently linked with a serine integrase, with or without a reverse transcriptase) or a CRISPR-

directed recombinase (e.g., CRISPR-Cas9 nickase covalently linked with a serine recombinase, with or without to a reverse transcriptase).

**[0154]** In some embodiments, the EDV comprises a gag protein fragment that is covalently linked to the polynucleotide-guided nuclease as a single polypeptide (i.e., a fusion protein).

**[0155]** In some embodiments, the EDV also comprises a targeting polypeptide that enables binding of the EDV to a target cell. In some embodiments, the EDV targeting polypeptide is fused to a transmembrane domain that anchors the targeting polypeptide in the EDV, enabling binding of the EDV to a target cell.

**[0156]** In some embodiments, the EDV also comprises a guide polynucleotide. In some embodiments, the guide polynucleotide forms a complex with the polynucleotide-guided nuclease. In some embodiments, the gag protein fragment and the polynucleotide-guided nuclease are covalently linked as a single polypeptide (i.e., a fusion protein).

**[0157]** In some cases, EDVs may also be used to deliver a polynucleotide that comprises a coding sequence to a target cell. For example, an EDV may comprise a coding sequence for a gag protein fragment, a viral envelope protein, a polynucleotide-guided nuclease, and/or a targeting polypeptide. In some cases, the EDV may comprise a coding sequence for a guide polynucleotide.

### i. Group-Specific Antigen (Gag) Protein

**[0158]** EDVs comprise a fragment of a gag protein, which can be present as part of a fusion with another component such as a nuclease and/or a targeting protein. In general, a naturally-occurring gag protein comprises (1) a matrix (MA) polypeptide that binds to cell membranes and directs virions to the cell surface, (2) a capsid (CA) polypeptide that forms an inner shell of a virus, and (3) a nucleocapsid (NC) polypeptide that binds directly to virion genomic material (e.g., RNA). A gag precursor can be expressed from viral genomic RNA as a single polypeptide that is later proteolytically cleaved to produce an MA polypeptide, a capsid polypeptide, and an NC polypeptide that are each separate from the others. A naturally-occurring gag protein also contains a (1) p2 polypeptide between the CA and NC polypeptides that plays a role in ordered virus assembly and infectivity, (2) a p1 polypeptide between the NC polypeptide and a p6 polypeptide that may play a role in virus infectivity, and (3) a p6 polypeptide that plays a role in the late steps of viral replication. Gag proteins are discussed in detail in Olson, Erik D, and Karin Musier-

Forsyth. "Retroviral Gag protein-RNA interactions: Implications for specific genomic RNA packaging and virion assembly." Seminars in cell & developmental biology vol. 86 (2019): 129-139. doi:10.1016/j.semcdb.2018.03.015; Hill, Melissa K et al. "Proline residues within spacer peptide p1 are important for human immunodeficiency virus type 1 infectivity, protein processing, and genomic RNA dimer stability." Journal of virology vol. 76,22 (2002): 11245-53. doi:10.1128/jvi.76.22.11245-11253.2002; Schmalen, Adrian et al. "The N-Terminus of the HIV-1 p6 Gag Protein Regulates Susceptibility to Degradation by IDE." Viruses vol. 10,12 710. 12 Dec. 2018, doi:10.3390/v10120710; Newman, John L et al. "Flexibility in the P2 domain of the HIV-1 Gag polyprotein." Protein science : a publication of the Protein Society vol. 13,8 (2004): 2101-7. doi:10.1110/ps.04614804.

**[0159]** An EDV gag protein fragment can be derived from a lentiviral gag protein or a retroviral gag protein. For example, the lentiviral gag protein can be selected from the group consisting of a bovine immunodeficiency virus gag protein, a simian immunodeficiency virus gag protein, a feline immunodeficiency virus gag protein, a human immunodeficiency virus (HIV) gag protein, an equine infection anemia virus gag protein, and a caprine arthritis encephalitis virus gag protein. In some embodiments, the gag protein is derived from an HIV gag protein.

**[0160]** The gag protein fragment in some embodiments can comprise an MA polypeptide and a CA polypeptide, and optionally does not compromise the NC polypeptide. The gag protein fragment in some embodiments can also comprise a p1 polypeptide, a p2 polypeptide, and/or a p6 polypeptide. The MA polypeptide, CA polypeptide, p1 polypeptide, p2 polypeptide, and p6 polypeptide can be derived from any lentiviral gag protein or retroviral gag protein. The polypeptides of the gag protein fragment can each be of a different origin and the polypeptides may be combined in a single synthetic gag protein fragment. In some embodiments, the gag protein fragment polypeptides are derived from an HIV gag protein

### 1. Matrix (MA) Polypeptide

**[0161]** Many MA polypeptides are suitable for use as an EDV MA polypeptide of the present disclosure. In many embodiments, the MA polypeptide is derived from a WT MA polypeptide, and comprises 30-135 amino acids (e.g., 30-131, 30-60, 30-40, 30-35, 33-135, 33-131, 33-60, 33-40, 33-35, 50-135, 50-131, 50-60, 55-57, 100-135, or 100-131) or all amino acids that correspond to the full-length sequence of a WT MA polypeptide.

**[0162]** In some embodiments, an MA polypeptide comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, or 100% identity with the amino acid sequence as set forth in SEQ ID NO: 59 as shown below.

GARASVLSGGELDRWEKIRLRPGGKKKYKLKHIVWASRELERFAVNPGLLETSEGCRQI LGQLQPSLQTGSEELRSLYNTVATLYCVHQRIEIKDTKEALDKIEEEQNKSKKKAQQAA ADTGHSNQVSQNY (SEQ ID NO: 59)

**[0163]** In some embodiments, the MA polypeptide is variant derived from the WT MA polypeptide, e.g., SEQ ID NO: 59. In these embodiments, the MA polypeptide is missing (e.g., via deletion) amino acids corresponding to amino acids at positions 72-127 of the sequence set forth in SEQ ID NO: 59. In these embodiments, the MA polypeptide is missing (e.g., via deletion) amino acids corresponding to amino acids at positions 30-127 of the sequence set forth in SEQ ID NO: 59. In these embodiments, the MA polypeptide is missing (e.g., via deletion) amino acids corresponding to amino acids at positions 53-127 of the sequence set forth in SEQ ID NO: 59.

**[0164]** In these embodiments, the MA polypeptide comprises a sequence that has at least 80% at least 85%, at least 90%, at least 95%, at least 97%, at least 98%, at least 99%, or 100% identity with the amino acid sequence as set forth in SEQ ID NO: 60 as shown below.

GARASVLSGGELDRWEKIRLRPGGKKKYKLKHIVWASRELERFAVNPGLLETSEGCRQI LGQLQPSLQTGSSQNY (SEQ ID NO: 60)

**[0165]** In these embodiments, the MA polypeptide comprises a sequence that has at least 80% at least 85%, at least 90%, at least 95%, at least 97%, at least 98%, at least 99%, or 100% identity with the amino acid sequence as set forth in SEQ ID NO: 61 as shown below.

GARASVLSGGELDRWEKIRLRPGGKKKYKLKHIVWASRELERFAVNPGLLETSQNY (SEQ ID NO: 61)

**[0166]** In these embodiments, the MA polypeptide comprises a sequence that has at least 80% at least 85%, at least 90%, at least 95%, at least 97%, at least 98%, at least 99%, or 100% identity with the amino acid sequence as set forth in SEQ ID NO: 62 as shown below.

GARASVLSGGELDRWEKIRLRPGGKKKYKSQNY (SEQ ID NO: 62)

### 2. Capsid (CA) Polypeptide

**[0167]** Many CA polypeptides are suitable for use as an EDV MA polypeptide of the present disclosure. In general, a CA polypeptide comprises an N-terminal portion and a C-terminal portion. In some embodiments, an EDV CA polypeptide comprises an amino acid sequence that corresponds to the full-length sequence of a WT CA polypeptide. In some embodiments, the EDV CA polypeptide comprises a truncation in the N-terminal portion, the C-terminal portion, or both. In many embodiments, the CA polypeptide is derived from a WT CA polypeptide, and comprises 80-220 amino acids (e.g., 80-210, 80-190, 80-180, 80-185, 80-175, 80-90, 170-220, 170-210, 170-190, 170-180, 170-185, 170-175, 170-90, 180-220, 180-210, 180-190, 180-180, 180-185, 180-175, 180-90, 195-220, 195-210, 195-190, 195-180, 195-185, 195-175, 195-90, 210-220, 210-210, 210-190, 210-180, 210-185, 210-175, 210-90, 83-220, 83-210, 83-190, 83-180, 83-185, 83-175, 83-90, 87-220, 87-210, 87-190, 87-180, 87-185, 87-175, 87-90, 174-220, 174-210, 174-190, 174-180, 174-185, 174-175, 174-90, 184-220, 184-210, or 184-190) or all amino acids that correspond to the full-length sequence of a WT CA polypeptide.

**[0168]** In some embodiments, the CA polypeptide comprises an amino acid sequences that corresponds to the full-length sequence of a WT CA protein. In some embodiments, the CA polypeptide comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, or 100% identity with the amino acid sequence as set forth in SEQ ID NO: 63 as shown below.

PIVQNIQGQMVHQAISPRTLNAWVKVVEEKAFSPEVIPMFSALSEGATPQDLNTMLNTV
GGHQAAMQMLKETINEEAAEWDRVHPVHAGPIAPGQMREPRGSDIAGTTSTLQEQIGW
MTHNPPIPVGEIYKRWIILGLNKIVRMYSPTSILDIRQGPKEPFRDYVDRFYKTLRAEQAS
QEVKNWMTETLLVQNANPDCKTILKALGPGATLEEMMTACQGVGGPGHKARVL (SEQ
ID NO: 63)

**[0169]** In some embodiments, the CA polypeptide comprises a sequence corresponding to amino acids at positions 149-231 of a WT CA polypeptide. In some embodiments, the CA polypeptide comprises a sequence corresponding to amino acids at positions 149-231 of a (SEQ ID NO: 63)

**[0170]** In some embodiments, the CA polypeptide comprises a sequence that lacks at least amino acids (i.e., has a deletion) corresponding to amino acids 5-148 of SEQ ID NO: 63. In some embodiments, the CA polypeptide comprises a sequence that lacks at least amino acids (i.e., has a

deletion) corresponding to amino acids 5-61 SEQ ID NO: 63. In some embodiments, the CA polypeptide comprises a sequence that lacks at least amino acids (i.e., has a deletion) corresponding to amino acids 5-47 SEQ ID NO: 63. In some embodiments, the CA polypeptide comprises a sequence that lacks at least amino acids (i.e., has a deletion) corresponding to amino acids 5-34 of SEQ ID NO: 63. In some embodiments, the CA polypeptide comprises a sequence that lacks at least amino acids (i.e., has a deletion) corresponding to amino acids 5-15 of SEQ ID NO: 63. In some embodiments, the CA polypeptide comprises a sequence that lacks at least amino acids (i.e., has a deletion) corresponding to amino acids 5-15 of SEQ ID NO: 63.

**[0171]** In some embodiments, the CA polypeptide comprises a sequence that has at least 80% at least 85%, at least 90%, at least 95%, at least 97%, at least 98%, at least 99%, or 100% identity with the amino acid sequence as set forth in SEQ ID NO: 64 as shown below.

PIVQSILDIRQGPKEPFRDYVDRFYKTLRAEQASQEVKNWMTETLLVQNANPDCKTILK ALGPGATLEEMMTACQGVGGPGHKARVL (SEQ ID NO: 64)

**[0172]** In some embodiments, the CA polypeptide comprises a sequence that has at least 80% at least 85%, at least 90%, at least 95%, at least 97%, at least 98%, at least 99%, or 100% identity with the amino acid sequence as set forth in SEQ ID NO: 65 as shown below.

PIVQHQAAMQMLKETINEEAAEWDRVHPVHAGPIAPGQMREPRGSDIAGTTSTLQEQIG WMTHNPPIPVGEIYKRWIILGLNKIVRMYSPTSILDIRQGPKEPFRDYVDRFYKTLRAEQ ASQEVKNWMTETLLVQNANPDCKTILKALGPGATLEEMMTACQGVGGPGHKARVL (SEQ ID NO: 65)

**[0173]** In some embodiments, the CA polypeptide comprises a sequence that has at least 80% at least 85%, at least 90%, at least 95%, at least 97%, at least 98%, at least 99%, or 100% identity with the amino acid sequence as set forth in SEQ ID NO: 66 as shown below.

HQAAMQMLKETINEEAAEWDRVHPVHAGPIAPGQMREPRGSDIAGTTSTLQEQIGWMT HNPPIPVGEIYKRWIILGLNKIVRMYSPTSILDIRQGPKEPFRDYVDRFYKTLRAEQASQE VKNWMTETLLVQNANPDCKTILKALGPGATLEEMMTACQGVGGPGHKARVL (SEQ ID NO: 66)

**[0174]** In some embodiments, the CA polypeptide comprises a sequence that has at least 80% at least 85%, at least 90%, at least 95%, at least 97%, at least 98%, at least 99%, or 100% identity with the amino acid sequence as set forth in SEQ ID NO: 67 as shown below.

PIVQTPQDLNTMLNTVGGHQAAMQMLKETINEEAAEWDRVHPVHAGPIAPGQMREPR GSDIAGTTSTLQEQIGWMTHNPPIPVGEIYKRWIILGLNKIVRMYSPTSILDIRQGPKEPFR

DYVDRFYKTLRAEQASQEVKNWMTETLLVQNANPDCKTILKALGPGATLEEMMTACQ
GVGGPGHKARVL (SEQ ID NO: 67)

**[0175]** In some embodiments, the CA polypeptide comprises a sequence that has at least 80% at least 85%, at least 90%, at least 95%, at least 97%, at least 98%, at least 99%, or 100% identity with the amino acid sequence as set forth in SEQ ID NO: 68 as shown below.

TPQDLNTMLNTVGGHQAAMQMLKETINEEAAEWDRVHPVHAGPIAPGQMREPRGSDI
AGTTSTLQEQIGWMTHNPPIPVGEIYKRWIILGLNKIVRMYSPTSILDIRQGPKEPFRDYV
DRFYKTLRAEQASQEVKNWMTETLLVQNANPDCKTILKALGPGATLEEMMTACQGVG
GPGHKARVL (SEQ ID NO: 68)

**[0176]** In some embodiments, the CA polypeptide comprises a sequence that has at least 80% at least 85%, at least 90%, at least 95%, at least 97%, at least 98%, at least 99%, or 100% identity with the amino acid sequence as set forth in SEQ ID NO: 69 as shown below.

PIVQEVIPMFSALSEGATPQDLNTMLNTVGGHQAAMQMLKETINEEAAEWDRVHPVHA
GPIAPGQMREPRGSDIAGTTSTLQEQIGWMTHNPPIPVGEIYKRWIILGLNKIVRMYSPTSI
LDIRQGPKEPFRDYVDRFYKTLRAEQASQEVKNWMTETLLVQNANPDCKTILKALGPG
ATLEEMMTACQGVGGPGHKARVL (SEQ ID NO: 69)

**[0177]** In some embodiments, the CA polypeptide comprises a sequence that has at least 80% at least 85%, at least 90%, at least 95%, at least 97%, at least 98%, at least 99%, or 100% identity with the amino acid sequence as set forth in SEQ ID NO: 70 as shown below.

EVIPMFSALSEGATPQDLNTMLNTVGGHQAAMQMLKETINEEAAEWDRVHPVHAGPIA
PGQMREPRGSDIAGTTSTLQEQIGWMTHNPPIPVGEIYKRWIILGLNKIVRMYSPTSILDIR
QGPKEPFRDYVDRFYKTLRAEQASQEVKNWMTETLLVQNANPDCKTILKALGPGATLE
EMMTACQGVGGPGHKARVL (SEQ ID NO: 70)

**[0178]** In some embodiments, the CA polypeptide comprises a sequence that has at least 80% at least 85%, at least 90%, at least 95%, at least 97%, at least 98%, at least 99%, or 100% identity with the amino acid sequence as set forth in SEQ ID NO: 71 as shown below.

PIVQSPRTLNAWVKVVEEKAFSPEVIPMFSALSEGATPQDLNTMLNTVGGHQAAMQML
KETINEEAAEWDRVHPVHAGPIAPGQMREPRGSDIAGTTSTLQEQIGWMTHNPPIPVGEI
YKRWIILGLNKIVRMYSPTSILDIRQGPKEPFRDYVDRFYKTLRAEQASQEVKNWMTET
LLVQNANPDCKTILKALGPGATLEEMMTACQGVGGPGHKARVL (SEQ ID NO: 71)

**[0179]** In some embodiments, the CA polypeptide comprises a sequence that has at least 80% at least 85%, at least 90%, at least 95%, at least 97%, at least 98%, at least 99%, or 100% identity with the amino acid sequence as set forth in SEQ ID NO: 72 as shown below.

SPRTLNAWVKVVEEKAFSPEVIPMFSALSEGATPQDLNTMLNTVGGHQAAMQMLKETI
NEEAAEWDRVHPVHAGPIAPGQMREPRGSDIAGTTSTLQEQIGWMTHNPPIPVGEIYKR
WIILGLNKIVRMYSPTSILDIRQGPKEPFRDYVDRFYKTLRAEQASQEVKNWMTETLLVQ
NANPDCKTILKALGPGATLEEMMTACQGVGGPGHKARVL (SEQ ID NO: 72)

### 3. Nucleoprotein (NC) Polypeptide

**[0180]** Many NC polypeptides are suitable for use as an EDV NC polypeptide of the present disclosure. In some embodiments, the NC polypeptide is derived from a WT MA polypeptide and comprises an amino acid sequence that corresponds to the full-length sequence or a truncated sequence of a WT NC polypeptide. In some embodiments, the EDV does not comprise an NC polypeptide.

**[0181]** In some embodiments, the gag protein fragment lacks an NC polypeptide. In some embodiments, instead of an NC polypeptide, the gag protein fragment comprises the sequence IQKGRQAN (SEQ ID NO: 73) instead.

**[0182]** In some embodiments, the NC polypeptide comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, or 100% identity with the amino acid sequence as set forth in SEQ ID NO: 74 as shown below.

IQKGNFRNQRKTVKCFNCGKEGHIAKNCRAPRKKGCWKCGKEGHQMKDCTERQAN
(SEQ ID NO: 74)

**[0183]** In some embodiments, the NC polypeptide is covalently linked with a p1-p6 polypeptide sequence as set forth in SEQ ID NO: 75 as shown below. In some embodiments, the p1-p6 polypeptide sequence is on the N-terminus of the EDV NC polypeptide. In some embodiments, the p1-p6 polypeptide sequence is on the C-terminus of the EDV NC polypeptide.

FLGKIWPSHKGRPGNFLQSRPEPTAPPEESFRFGEETTTPSQKQEPIDKELYPLASLRSLFG
SDPSSQ (SEQ ID NO: 75)

**[0184]** In some embodiments, the NC-p1-p6 polypeptide comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, or 100% identity with the amino acid sequence as set forth in SEQ ID NO: 76 as shown below.

IQKGNFRNQRKTVKCFNCGKEGHIAKNCRAPRKKGCWKCGKEGHQMKDCTERQANFL GKIWPSHKGRPGNFLQSRPEPTAPPEESFRFGEETTTPSQKQEPIDKELYPLASLRSLFGSD PSSQ (SEQ ID NO: 76)

**[0185]** In some embodiments, the NC polypeptide is covalently linked with a p2 polypeptide sequence as set forth in SEQ ID NO: 77 as shown below. In some embodiments, the p2 polypeptide sequence is on the N-terminus of the EDV NC polypeptide. In some embodiments, the p2 polypeptide sequence is on the C-terminus of the EDV NC polypeptide.

AEAMSQVTNPATIM (SEQ ID NO: 77)

### 4. Protease Cleavage Sites

**[0186]** In some embodiments, the gag protein fragment and the polynucleotide-guided nuclease are covalently linked as a single polypeptide (i.e., a fusion protein). In some embodiments, the fusion protein optionally includes one or more heterologous protease cleavage sites between the gag protein fragment and the polynucleotide-guided nuclease (e.g., CRISPR-Cas nuclease, CAST, CRISPR-directed integrase, CRISPR-directed recombinase, or variant thereof). Presence of at least one cleavage site enables the release of the polynucleotide-guided nuclease from the gag protein fragment so that the EDV can properly assemble and mature in a production cell line. Many protease cleavage sites will be known to one of ordinary skill in the art, and any convenient protease cleavage site can be used. In some embodiments, the heterologous protease cleavage site is a TEV cleavage site, a PreScission cleavage site, a human rhinovirus 3C protease cleavage site, an enterokinase cleavage site, an Epstein-Barr virus protease cleavage site, a cathepsin D cleavage site, and/or a thrombin cleavage site, or any combination thereof.

### ii. Viral Envelope Protein

**[0187]** The viral envelope protein of the EDV is typically a viral envelope glycoprotein. The viral envelope protein can catalyze lipid bilayer remodeling, causing fusion and/or fission of lipid bilayers, such as those found in cell membranes and endosomes. In some embodiments, the EDV viral envelope protein is a fusogenic variant of a WT viral envelope glycoprotein, wherein the variant is capable of causing more fusion and/or fission of lipid bilayers (i.e., more "fusogenic") compared to the WT protein from which it is derived. Suitable EDV viral envelope proteins include, e.g., a vesicular stomatitis virus (VSV) glycoprotein (VSVG protein), a Measles virus hemagglutinin (HA) protein and/or a measles virus fusion glycoprotein, an Influenza virus

neuraminidase (NA) protein, a Measles virus F protein, an Influenza virus HA protein, a Moloney virus MLV-A protein, a Moloney virus MLV-E protein, a Baboon Endogenous retrovirus (BAEV) envelope protein, an Ebola virus glycoprotein, a foamy virus envelope protein, and a combination of two or more of the foregoing viral envelope proteins. Suitable EDV viral envelope proteins also include fusogenic variants of any of the foregoing viral envelope proteins.

**[0188]** In some embodiments, the EDV viral envelope protein is a VSVG protein. In some embodiments, the viral envelope protein is a measles virus hemagglutinin protein. In some embodiments, the viral envelope protein is a measles virus F protein. In some embodiments, the viral envelope protein is an influenza virus hemagglutinin protein. In some embodiments, the viral envelope protein is a Moloney virus MLV-A protein. In some embodiments, the viral envelope protein is a Moloney virus MLV-E protein. In some embodiments, the viral envelope protein is a baboon endogenous retrovirus envelope protein. In some embodiments, the viral envelope protein is an Ebola virus glycoprotein. In some embodiments, the viral envelope protein is a foamy virus envelope protein.

**[0189]** In some embodiments, the EDV viral envelope protein is a VSVG protein. In some embodiments, a suitable VSVG protein comprises an amino acid sequence having at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, or 100% amino acid sequence identity to the amino acid sequence as set forth in SEQ ID NO: 78-80 and provided below.

**[0190]** In some embodiments, the VSVG protein comprises an amino acid sequence having at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, or 100% amino acid sequence identity to the amino acid sequence as set forth in SEQ ID NO: 80, and the amino acid at position 47 of SEQ ID NO: 80 is other than a Lys. In some embodiments, the VSVG protein comprises an amino acid sequence having at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, or 100% amino acid sequence identity to the amino acid sequence as set forth in SEQ ID NO: 80, and the amino acid at position 354 of SEQ ID NO: 80 is other than an Arg. In some embodiments, the VSVG protein comprises an amino acid sequence having at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, or 100% amino acid sequence identity to the amino acid sequence as set forth in SEQ ID NO: 80, the amino acid at position 47 of SEQ ID NO: 80 is other than a Lys, and the amino acid at position 354 of SEQ ID NO: 80 is other than an Arg. In some embodiments, the Lys at amino acid 47 of SEQ ID NO: 80

is substituted with an Ala. In some embodiments, the Lys at amino acid 47 of SEQ ID NO: 80 is substituted with a Gln. In some embodiments, the Arg at amino acid 354 of SEQ ID NO: 80 is substituted with an Ala. In some embodiments, the Arg at amino acid 354 of SEQ ID NO: 80 is substituted with a Gln. In some embodiments, the Lys at amino acid 47 of SEQ ID NO: 80 is substituted with an Gln, and the Arg at amino acid 354 of SEQ ID NO: 80 is substituted with an Ala.

**[0191]** In some embodiments, a suitable VSVG protein comprises an amino acid sequence having at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, or 100% amino acid sequence identity to the amino acid sequence as set forth in SEQ ID NO: 81 and provided below in Table 2. In some embodiments, the VSVG protein comprises an amino acid sequence having at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, or 100% amino acid sequence identity to the amino acid sequence as set forth in SEQ ID NO: 81, the amino acid at position 47 of SEQ ID NO: 81 is an Ala, and the amino acid at position 354 of SEQ ID NO: 81 is an Ala. In some embodiments, the VSVG protein comprises an amino acid sequence having at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, or 100% amino acid sequence identity to the amino acid sequence as set forth in SEQ ID NO: 81, the amino acid at position 47 of SEQ ID NO: 81 is a Gln and the amino acid at position 354 of SEQ ID NO: 81 is an Ala.

**Table 2 – VSVG and VSVGmut Sequences**

| SEQ ID NO | Description | Sequence |
|---|---|---|
| 78 | VSVG protein | IMKCLLYLAFLFIGVNCKFTIVFPHNQKGNWKNVPSNYH YCPSSSDLNWHNDLIGTALQVKMPKSHKAIQADGWMC HASKWVTTCDFRWYGPKYITHSIRSFTPSVEQCKESIEQT KQGTWLNPGFPPQSCGYATVTDAEAVIVQVTPHHVLVD EYTGEWVDSQFINGKCSNYICPTVHNSTTWHSDYKVKG LCDSNLISMDITFFSEDGELSSLGKEGTGFRSNYFAYETG GKACKMQYCKHWGVRLPSGVWFEMADKDLFAAARFP ECPEGSSISAPSQTSVDVSLIQDVERILDYSLCQETWSKIR AGLPISPVDLSYLAPKNPGTGPAFTIINGTLKYFETRYIRV DIAAPILSRMVGMISGTTTERELWDDWAPYEDVEIGPNG VLRTSSGYKFPLYMIGHGMLDSDLHLSSKAQVFEHPHIQ DAASQLPDDESLFFGDTGLSKNPIELVEGWFSSWKSSIAS FFFIIGLIIGLFLVLRVGIHLCIKLKHTKKRQIYTDIEMNRL GK |

| | | |
|---|---|---|
| 79 | VSVG protein | MKCLLYLAFLFIGVNCKFTIVFPHNQKGNWKNVPSNYH YCPSSSDLNWHNDLIGTALQVKMPKSHKAIQADGWMC HASKWVTTCDFRWYGPKYITHSIRSFTPSVEQCKESIEQT KQGTWLNPGFPPQSCGYATVTDAEAVIVQVTPHHVLVD EYTGEWVDSQFINGKCSNYICPTVHNSTTWHSDYKVKG LCDSNLISMDITFFSEDGELSSLGKEGTGFRSNYFAYETG GKACKMQYCKHWGVRLPSGVWFEMADKDLFAAARFP ECPEGSSISAPSQTSVDVSLIQDVERILDYSLCQETWSKIR AGLPISPVDLSYLAPKNPGTGPAFTIINGTLKYFETRYIRV DIAAPILSRMVGMISGTTTERELWDDWAPYEDVEIGPNG VLRTSSGYKFPLYMIGHGMLDSDLHLSSKAQVFEHPHIQ DAASQLPDDESLFFGDTGLSKNPIELVEGWFSSWKSSIAS FFFIIGLIIGLFLVLRVGIHLCIKLKHTKKRQIYTDIEMNRL GK |
| 80 | VSVG protein | KFTIVFPHNQKGNWKNVPSNYHYCPSSSDLNWHNDLIG TALQVKMPKSHKAIQADGWMCHASKWVTTCDFRWYG PKYITHSIRSFTPSVEQCKESIEQTKQGTWLNPGFPPQSCG YATVTDAEAVIVQVTPHHVLVDEYTGEWVDSQFINGKC SNYICPTVHNSTTWHSDYKVKGLCDSNLISMDITFFSED GELSSLGKEGTGFRSNYFAYETGGKACKMQYCKHWGV RLPSGVWFEMADKDLFAAARFPECPEGSSISAPSQTSVD VSLIQDVERILDYSLCQETWSKIRAGLPISPVDLSYLAPK NPGTGPAFTIINGTLKYFETRYIRVDIAAPILSRMVGMISG TTTERELWDDWAPYEDVEIGPNGVLRTSSGYKFPLYMIG HGMLDSDLHLSSKAQVFEHPHIQDAASQLPDDESLFFGD TGLSKNPIELVEGWFSSWKSSIASFFFIIGLIIGLFLVLRVG IHLCIKLKHTKKRQIYTDIEMNRLGK |
| 81 | VSVGmut; underlined and bold letters indicate K47Q and R354A mutations | KFTIVFPHNQKGNWKNVPSNYHYCPSSSDLNWHNDLIG TALQVQMP**Q**SHKAIQADGWMCHASKWVTTCDFRWYG PKYITHSIRSFTPSVEQCKESIEQTKQGTWLNPGFPPQSCG YATVTDAEAVIVQVTPHHVLVDEYTGEWVDSQFINGKC SNYICPTVHNSTTWHSDYKVKGLCDSNLISMDITFFSED GELSSLGKEGTGFRSNYFAYETGGKACKMQYCKHWGV RLPSGVWFEMADKDLFAAARFPECPEGSSISAPSQTSVD VSLIQDVERILDYSLCQETWSKIRAGLPISPVDLSYLAPK NPGTGPAFTIINGTLKYFETRYIRVDIAAPILSRMVGMISG TTTE**A**ELWDDWAPYEDVEIGPNGVLRTSSGYKFPLYMI GHGMLDSDLHLSSKAQVFEHPHIQDAASQLPDDESLFFG DTGLSKNPIELVEGWFSSWKSSIASFFFIIGLIIGLFLVLRV GIHLCIKLKHTKKRQIYTDIEMNRLGK |

**[0192]** In some embodiments, the EDV viral envelope protein is an Ebola Zaire virus glycoprotein. In some embodiments, the Ebola Zaire virus glycoprotein comprises an amino acid

sequence having at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, or 100% amino acid sequence identity with the following amino acid sequence as set forth in SEQ ID NO: 82 (GenBank Accession No: AAB81004).

MGVTGILQLPRDRFKRTSFFLWVIILFQRTFSIPLGVIHNSTLQVSDVDKLVCRDKLSSTN
QLRSVGLNLEGNGVATDVPSATKRWGFRSGVPPKVVNYEAGEWAENCYNLEIKKPDG
SECLPAAPDGIRGFPRCRYVHKVSGTGPCAGDFAFHKEGAFFLYDRLASTVIYRGTTFAE
GVVAFLILPQAKKDFFSSHPLREPVNATEDPSSGYYSTTIRYQATGFGTNETEYLFEVDN
LTYVQLESRFTPQFLLQLNETIYTSGKRSNTTGKLIWKVNPEIDTTIGEWAFWETKKNLT
RKIRSEELSFTVVSNGAKNISGQSPARTSSDPGTNTTTEDHKIMASENSSAMVQVHSQGR
EAAVSHLTTLATISTSPQSLTTKPGPDNSTHNTPVYKLDISEATQVEQHHRRTDNDSTAS
DTPSATTAAGPPKAENTNTSKSTDFLDPATTTSPQNHSETAGNNNTHHQDTGEESASSG
KLGLITNTIAGVAGLITGGRRTRREAIVNAQPKCNPNLHYWTTQDEGAAIGLAWIPYFGP
AAEGIYIEGLMHNQDGLICGLRQLANETTQALQLFLRATTELRTFSILNRKAIDFLLQRW
GGTCHILGPDCCIEPHDWTKNITDKIDQIIHDFVDKTLPDQGDNDNWWTGWRQWIPAGI
GVTGVIIAVIALFCICKFVF (SEQ ID NO: 82)

**[0193]** In some embodiments, the EDV viral envelope protein is an Ebola Zaire virus glycoprotein. In some embodiments, the Ebola Zaire virus glycoprotein comprises an amino acid sequence having at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, or 100% amino acid sequence identity with the following amino acid sequence as set forth in SEQ ID NO: 83.

IPLGVIHNSTLQVSDVDKLVCRDKLSSTNQLRSVGLNLEGNGVATDVPSATKRWGFRSG
VPPKVVNYEAGEWAENCYNLEIKKPDGSECLPAAPDGIRGFPRCRYVHKVSGTGPCAG
DFAFHKEGAFFLYDRLASTVIYRGTTFAEGVVAFLILPQAKKDFFSSHPLREPVNATEDP
SSGYYSTTIRYQATGFGTNETEYLFEVDNLTYVQLESRFTPQFLLQLNETIYTSGKRSNTT
GKLIWKVNPEIDTTIGEWAFWETKKNLTRKIRSEELSFTVVSNGAKNISGQSPARTSSDP
GTNTTTEDHKIMASENSSAMVQVHSQGREAAVSHLTTLATISTSPQSLTTKPGPDNSTHN
TPVYKLDISEATQVEQHHRRTDNDSTASDTPSATTAAGPPKAENTNTSKSTDFLDPATTT
SPQNHSETAGNNNTHHQDTGEESASSGKLGLITNTIAGVAGLITGGRRTRREAIVNAQPK
CNPNLHYWTTQDEGAAIGLAWIPYFGPAAEGIYIEGLMHNQDGLICGLRQLANETTQAL
QLFLRATTELRTFSILNRKAIDFLLQRWGGTCHILGPDCCIEPHDWTKNITDKIDQIIHDFV
DKTLPDQGDNDNWWTGWRQWIPAGIGVTG VIIAVIALFCICKFVF (SEQ ID NO: 83)

**[0194]** In some embodiments, the EDV viral envelope protein is an Ebola Reston virus glycoprotein. In some embodiments, the Ebola Reston virus glycoprotein an amino acid sequence having at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, or 100% amino acid sequence identity with the following amino acid sequence as set forth in SEQ ID NO: 84 (GenBank Accession No: NP_690583).

MGSGYQLLQLPRERFRKTSFLVWVIILFQRAISMPLGIVTNSTLKATEIDQLVCRDKLSS
TSQLKSVGLNLEGNGIATDVPSATKRWGFRSGVPPKVVSYEAGEWAENCYNLEIKKSD
GSECLPLPPDGVRGFPRCRYVHKVQGTGPCPGDLAFHKNGAFFLYDRLASTVIYRGTTF
AEGVVAFLILSEPKKHFWKATPAHEPVNTTDDSTSYYMTLTLSYEMSNFGGNESNTLFK
VDNHTYVQLDRPHTPQFLVQLNETLRRNNRLSNSTGRLTWTLDPKIEPDVGEWAFWET
KKNFSQQLHGENLHFQIPSTHTNNSSDQSPAGTVQGKISYHPPANNSELVPTDSPPVVSV
LTAGRTEEMSTQGLTNGETITGFTANPMTTTIAPSPTMTSEVDNNVPSEQPNNTASIEDSP
PSASNETIYHSEMDPIQGSNNSAQSPQTKTTPAPTTSPMTQDPQETANSSKPGTSPGSAAG
PSQPGLTINTVSKVADSLSPTRKQKRSVRQNTANKCNPDLYYWTAVDEGAAVGLAWIP
YFGPAAEGIYIEGVMHNQNGLICGLRQLANETTQALQLFLRATTELRTYSLLNRKAIDFL
LQRWGGTCRILGPSCCIEPHDWTKNITDEINQIKHDFIDNPLPDHGDDLNLWTGWRQWIP
AGIGIIGVIIAIIALLCICKILC (SEQ ID NO: 84)

**[0195]** In some embodiments, the EDV viral envelope protein is a Marburg virus glycoprotein. In some embodiments, the Marburg virus glycoprotein comprises an amino acid sequence having at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, or 100% amino acid sequence identity with the following amino acid sequence as set forth in SEQ ID NO: 85 (GenBank Accession No: CAA78117).

MKTTCFLISLILIQGTKNLPILEIASNNQPQNVDSVCSGTLQKTEDVHLMGFTLSGQKVA
DSPLEASKRWAFRTGVPPKNVEYTEGEEAKTCYNISVTDPSGKSLLLDPPTNIRDYPKCK
TIHHIQGQNPHAQGIALHLWGAFFLYDRIASTTMYRGKVFTEGNIAAMIVNKTVHKMIF
SRQGQGYRHMNLTSTNKYWTSSNGTQTNDTGCFGALQEYNSTKNQTCAPSKIPPPLPTA
RPEIKLTSTPTDATKLNTTDPSSDDEDLATSGSGSGEREPHTTSDAVTKQGLSSTMPPTPS
PQPSTPQQGGNNTNHSQDAVTELDKNNTTAQPSMPPHNTTTISTNNTSKHNFSTLSAPLQ
NTTNDNTQSTITENEQTSAPSITTLPPTGNPTTAKSTSSKKGPATTAPNTTNEHFTSPPPTP
SSTAQHLVYFRRKRSILWREGDMFPFLDGLINAPIDFDPVPNTKTIFDESSSSGASAEEDQ
HASPNISLTLSYFPNINENTAYSGENENDCDAELRIWSVQEDDLAAGLSWIPFFGPGIEGL
YTAVLIKNQNNLVCRLRRLANQTAKSLELLLRVTTEERTFSLINRHAIDFLLTRWGGTCK
VLGPDCCIGIEDLSKNISEQIDQIKKDEQKEGTGWGLGGKWWTSDWGVLTNLGILLLLSI
AVLIALSCICRIFTKYIG (SEQ ID NO: 85)

**[0196]** In some embodiments, the EDV viral envelope protein is a Venezuelan equine encephalitis virus glycoprotein. In some embodiments, the Venezuelan equine encephalitis virus glycoprotein comprises an amino acid sequence having at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, or 100% amino acid sequence identity with the following amino acid sequence as set forth in SEQ ID NO: 86 (GenBank Accession No: AAU89534).

MFPFQPMYPMQPMPYRNPFAAPRRPWFPRTDPFLAMQVQELTRSMANLTFKQRRDAPP
EGPSAKKPKKEASQKQKGGGQGKKKKNQGKKKAKTGPPNPKAQNGNKKKTNKKPGK
RQRMVMKLESDKTFPIMLEGKINGYACVVGGKLFRPMHVEGKIDNDVLAALKTKKAS
KYDLEYADVPQNMRADTFKYTHEKPQGYYSWHHGAVQYENGRFTVPKGVGAKGDSG
RPILDNQGRVVAIVLGGVNEGSRTALSVVMWNEKGVTVKYTPENCEQWSLVTTMCLL

ANVTFPCAQPPICYDRKPAETLAMLSVNVDNPGYDELLEAAVKCPGSTEELFKEYKLTR
PYMARCIRCAVGSCHSPIAIEAVKSDGHDGYVRLQTSSQYGLDSSGNLKGRTMRYDM
HGTIKEIPLHQVSLHTSRPCHIVDGHGYFLLARCPAGDSITMEFKKDSVTHSCSVPYEVK
FNPVGRELYTHPPEHGVEQACQVYAHDAQNRGAYVEMHLPGSEVDSSLVSLSGSSVTV
TPPVGTSALVECECGGTKISETINKTKQFSQCTKKEQCRAYRLQNDKWVYISDKLPKAA
GATLKGKLHVPFLLADGKCTVPLAPEPMITFGFRSVSLKLHPKNPTYLTTRQLADEPHYT
HELISEPAVRNFTVTGKGWEFVWGNHPPKRFWAQETAPGNPHGLPHEVITHYYHRYPM
STILGLSICAAIATVSVAASTWLFCRSRVACLTPYRLTPNARIPFCLAVLCCARTARAETT
WESLDHLWNNNQQMFWIQLLIPLAALIVVTRLLRCVCCVVPFLVMAGAAGAGAYEHA
TTMPSQAGISYNTIVNRAGYAPLPISITPTKIKLIPTVNLEYVTCHYKTGMDSPAIKCCGS
QECTPTYRPDEQCKVFTGVYPFMWGGAYCFCDTENTQVSKAYVMKSDDCLADHAEAY
KAHTASVQAFLNITVGEHSIVTTVYVNGETPVNFNGVKLTAGPLSTAWTPFDRKIVQYA
GEIYNYDFPEYGAGQPGAFGDIQSRTVSSSDLYANTNLVLQRPKAGAIHVPYTQAPSGFE
QWKKDKAPSLKSTAPFGCEIYTNPIRAENCAVGSIPLAFDIPDALFTRVSETPTLSAAECT
LNECVYSSDFGGIATVKYSASKSGKCAVHVPSGTATLKEAAVELTEQGSATIHFSTANIH
PEFRLQICTSYVTCKGDCHPPKDHIVTHPQYHAQTFTAAVSKTAWTWLTSLLGGSAVIIII
GLVLATIVAMYVLTNQKHN (SEQ ID NO: 86)

**[0197]** In some embodiments, the EDV viral envelope protein is a Venezuelan equine encephalitis virus E2 glycoprotein. In some embodiments, the Venezuelan equine encephalitis virus E2 glycoprotein comprises an amino acid sequence having at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, or 100% amino acid sequence identity with the following amino acid sequence as set forth in SEQ ID NO: 87 (GenBank Accession No: AAU89534).

STEELFKEYKLTRPYMARCIRCAVGSCHSPIAIEAVKSDGHDGYVRLQTSSQYGLDSSGN
LKGRTMRYDMHGTIKEIPLHQVSLHTSRPCHIVDGHGYFLLARCPAGDSITMEFKKDSV
THSCSVPYEVKFNPVGRELYTHPPEHGVEQACQVYAHDAQNRGAYVEMHLPGSEVDSS
LVSLSGSSVTVTPPVGTSALVECECGGTKISETINKTKQFSQCTKKEQCRAYRLQNDKW
VYISDKLPKAAGATLKGKLHVPFLLADGKCTVPLAPEPMITFGFRSVSLKLHPKNPTYLT
TRQLADEPHYTHELISEPAVRNFTVTGKGWEFVWGNHPPKRFWAQETAPGNPHGLPHE
VITHYYHRYPMSTILGLSICAAIATVSVAASTWLFCRSRVACLTPYRLTPNARIPFCLAVL
CCART ARA (SEQ ID NO: 87)

**[0198]** In some embodiments, the EDV viral envelope protein is a Venezuelan equine encephalitis virus E1 glycoprotein. In some embodiments, the Venezuelan equine encephalitis virus E1 glycoprotein comprises an amino acid sequence having at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, or 100% amino acid sequence identity with the following amino acid sequence as set forth in SEQ ID NO: 88 (GenBank Accession No: AAU89534).

YEHATTMPSQAGISYNTIVNRAGYAPLPISITPTKIKLIPTVNLEYVTCHYKTGMDSPAIK
CCGSQECTPTYRPDEQCKVFTGVYPFMWGGAYCFCDTENTQVSKAYVMKSDDCLADH
AEAYKAHTASVQAFLNITVGEHSIVTTVYVNGETPVNFNGVKLTAGPLSTAWTPFDRKI
VQYAGEIYNYDFPEYGAGQPGAFGDIQSRTVSSSDLYANTNLVLQRPKAGAIHVPYTQA
PSGFEQWKKDKAPSLKSTAPFGCEIYTNPIRAENCAVGSIPLAFDIPDALFTRVSETPTLSA
AECTLNECVYSSDFGGIATVKYSASKSGKCAVHVPSGTATLKEAAVELTEQGSATIHFST
ANIHPEFRLQICTSYVTCKGDCHPPKDHIVTHPQYHAQTFTAAVSKTAWTWLTSLLGGS
AVIIIIGLVLATIVAMYVLTNQKHN (SEQ ID NO: 88)

**[0199]** In some embodiments, the EDV viral envelope protein is a human T-lymphotropic virus 1 (HTLV-1) glycoprotein. In some embodiments, the human T-lymphotropic virus 1 (HTLV-1) glycoprotein comprises an amino acid sequence having at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, or 100% amino acid sequence identity with the following amino acid sequence as set forth in SEQ ID NO: 89 (GenBank Accession No: AAU04884).

MGKFLATLILFFQFCPLILGDYSPSCCTLTVGVSSYHSKPCNPAQPVCSWTLDLLALSAD
QALQPPCPNLVSYSSYHATYSLYLFPHWIKKPNRNGGGYYSASYSDPCSLKCPYLGCQS
WTCPYTGAVSSPYWKFQQDVNFTQEVSHLNINLHFSKCGFPFSLLVDAPGYDPIWFLNT
EPSQLPPTAPPLLSHSNLDHILEPSIPWKSKLLTLVQLTLQSTNYTCIVCIDRASLSTWHVL
YSPNVSVPSLSSTPLLYPSLALPAPHLTLPFNWTHCFDPQIQAIVSSPCHNSLILPPFSLSPV
PTLGSRSRRAVPVAVWLVSALAMGAGVAGGITGSMSLASGKSLLHEVDKDISQLTQA
IVKNHKNLLKIAQYAAQNRRGLDLLFWEQGGLCKALQEQCCFLNITNSHVSILQERPPLE
NRVLTGWGLNWDLGLSQWAREALQTGITLVALLLLVILAGPCILRQLRHLPSRVRYPHY
SLINPESSL (SEQ ID NO: 89)

**[0200]** In some embodiments, the EDV viral envelope protein is a measles virus hemagglutinin (H) polypeptide. In some embodiments, the measles virus hemagglutinin (H) polypeptide comprises an amino acid sequence having at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, or 100% amino acid sequence identity with the following amino acid sequence as set forth in SEQ ID NO: 90 (GenBank Accession No: AAU04884).

MSPQRDRINAFYKDNPHPKGSRIVINREHLMIDRPYVLLAVLFVMFLSLIGLLAIAGIRLH
RAAIYTAEIHKSLSTNLDVTNSIEHQVKDVLTPLFKIIGDEVGLRTPQRFTDLVKFISDKIK
FLNPDREYDFRDLTWCINPPERIKLDYDQYCADVAAEELMNALVNSTLLETRTTNQFLA
VSKGNCSGPTTIRGQFSNMSLSLLDLYLSRGYNVSSIVTMTSQGMYGGTYLVEKPNLSS
KGSELSQLSMYRVFEVGVIRNPGLGAPVFHMTNYFEQPVSNDLSNCMVALGELKLAAL
CHGGDSITIPYQGSGKGVSFQLVKLGVWKSPTDMQSWVPLSTDDPVIDRLYLSSHRGVI
ADNQAKWAVPTTRTDDKLRMETCFQQACKGKIQALCENPEWAPLKDNRIPSYGVLSVD
LSLTVELKIKIASGFGPLITHGSGMDLYKSNHNNVYWLTIPPMKNLALGVINTLEWIPRF
KVSPYLFTVPIKEAGEDCHAPTYLPAEVDGDVKLSSNLVILPGQDLQYVLATYDTSRVE
HAVVYYVYSPSRSFSYFYPFRLPIKGIPIELQVECFTWDQKLWCRHFCVLADSESGGHIT
HSGMVGMGVSCTVTREDGTNSR (SEQ ID NO: 90)

**[0201]** In some embodiments, the EDV viral envelope protein is a measles virus fusion (F) polypeptide. In some embodiments, the measles virus fusion (F) polypeptide comprises an amino acid sequence having at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, or 100% amino acid sequence identity with the following amino acid sequence as set forth in SEQ ID NO: 91 (GenBank Accession No: AAU04884).

MSIMGLKVNVSAIFMAVLLTLQTPTGQIHWGNLSKIGVVGIGSASYKVMTRSSHQSLVI
KLMPNITLLNNCTRVEIAEYRRLLRTVLEPIRDALNAMTQNIRPVQSVASSRRHKRFAGV
VLAGAALGVATAAQITAGIALHQSMLNSQAIDNLRASLETTNQAIETIRQAGQEMILAV
QGVQDYINNELIPSMNQLSCDLIGQKLGLKLLRYYTEILSLFGPSLRDPISAEISIQALSYA
LGGDINKVLEKLGYSGGDLLGILESGGIKARITHVDTESYFIVLSIAYPTLSEIKGVIVHRL
EGVSYNIGSQEWYTTVPKYVATQGYLISNFDESSCTFMPEGTVCSQNALYPMSPLLQE
CLRGYTKSCARTLVSGSFGNRFILSQGNLIANCASILCKCYTTGTIINQDPDKILTYIAA
DHCPVVEVNGVTIQVGSRRYPDAVYLHRIDLGPPISLERLDVGTNLGNAIAKLEDAKELL
ESSDQILRSMKGLSSTSIVYILIAVCLGGLIGIPALICCCRGRCNKKGEQVGMSRPGLKP
DLTGTSKSYVRSL (SEQ ID NO: 91)

### iii. Targeting Polypeptide

**[0202]** Targeting polypeptides of an EDV enable delivery of genome-editing molecules to specific target cells. A targeting polypeptide has specific binding for a molecule that is expressed on the surface of a target cell. By recognizing cell surface marker(s) on a target cell, delivery of genome-editing to molecules are improved. Thus, an EDV can bind specifically to a target cell, thereby delivering genome-editing molecules (e.g., polynucleotide-guided nucleases, with or without guide polynucleotides) to the target cell. In some embodiments, the targeting polypeptide is an antibody or an antibody fragment that binds to the molecule on the target cell, e.g., an scFv, a diabody, a triabody, a nanobody, and any antigen-binding fragment comprising the $V_H$ and $V_L$ domains of an antibody. In some embodiments, the targeting polypeptide is a bispecific antibody. In some embodiments, the targeting polypeptide is an antibody analog, e.g., an affibody, an affilin, an affimer, an affitin, an αbody, an anticalin, an avimer, a DARPin, a Fynomer, a Kunitz domain peptide, a monobody, a repebody, a VLR, and a nanoCLAMP. In some embodiments, the targeting polypeptide is a natural ligand of the molecule on the target cell, e.g., a cytokine.

**[0203]** An EDV may comprise one polypeptide with one, two, three, four, five, or more target-binding regions. An EDV may comprise one, two, three, four, five, or more targeting polypeptides. Target-binding regions or targeting polypeptides with different binding specificities may be combined and used together for delivering molecules to the same target cell or same target cell

type. For example, two vectors may each target a different molecule on the surface of T cells, e.g., CD3 and CD28, and the two vectors may be used together to deliver their molecules to the same T cells. In another example, two vectors may each target the same molecule on the surface of a cell, e.g., CD3 on a T cell, but each of the vectors uses a different mechanism of binding to the same molecule.

**[0204]** An EDV may also comprise a bispecific targeting polypeptide that binds to two different targets on the same target cell or same target cell type. For example, a bispecific targeting polypeptide bind to different molecules on the surface of T cells, e.g., CD3 and CD28.

**[0205]** In some embodiments, the target cell is an immune cell, such as a T cell, a B cell, a natural killer (NK) cell, a mast cell, a dendritic cell, a macrophage, a monocyte, and other immune cell types. In some embodiments, the target cell is a type of stem cell, such as an HSC.

**[0206]** In some embodiments, the EDV selectively targets T cells. In some embodiments, the targeting polypeptide binds to CD3, CD4, CD5, CD7, CD8, CD19, CD28, 4-1BB ligand, T cell receptor (TCR) α constant chain, TCR beta constant chain, or a major histocompatibility complex (MHC) carrying T cell receptor (TCR) specific peptide. In some embodiments, the targeting polypeptide comprises one or more antibodies, e.g., one or a combination of anti-CD3 (e.g., CD3 scFv-3), anti-CD4 (e.g., CD4 scFv-2), and anti-CD28 (e.g., CD28 scFv-2) antibody. In some embodiments, the targeting polypeptide comprises an anti-CD3 and an anti-CD4 antibody, an anti-CD3 and an anti-CD28 antibody, an anti-CD3, an anti-CD4, or an anti-CD28 antibody.

**[0207]** In some embodiments, the targeting polypeptide comprises an anti-CD3 antibody or fragment thereof, e.g., an anti-CD3 scFv or anti-CD3 Fab. In some embodiments, the anti-CD3 scFv or anti-CD3 Fab comprises the amino acid sequence as set forth in SEQ ID NO: 97. In some embodiments, the anti-CD3 scFv or anti-CD3 Fab comprises the heavy chain and light chain CDRs in a heterologous heavy chain and light chain framework region, respectively, or the entire heavy and light chain variable regions, of the OKT3 antibody (see, e.g., Kjer-Nielsen, et al. *PNAS USA* 101 (20) 7675-7680 (2004). The heavy chain variable region ($V_H$) of the OKT antibody, with CDRs underlined, is:

QVQLQQSGAELARPGASVKMSCKASGYTFTRYTMHWVKQRPGQGLEWIGYINPSRGY TNYNQKFKDKATLTTDKSSSTAYMQLSSLTSEDSAVYYCARYYDDHYCLDYWGQGTT LTVSS (SEQ ID NO: 264).

The light chain variable region ($V_L$) of the OKT antibody, with CDRs underlined, is:

QIVLTQSPAIMSASPGEKVTMTCSASSSVSYMNWYQQKSGTSPKRWIYDTSKLASGVPA HFRGSGSGTSYSLTISGMEAEDAATYYCQQWSSNPFTFGSGTKLEIN (SEQ ID NO: 265).

**[0208]** In some embodiments, the anti-CD3 scFv or anti-CD3 Fab comprises the heavy chain and light chain CDRs in a heterologous heavy chain and light chain framework region, respectively, or the entire heavy and light chain variable regions, of the Hit3a antibody (see, e.g., Chinese patent application number CN-1298020). The heavy chain variable region ($V_H$) of the Hit3a antibody, with CDRs underlined, is:

QVQLQESGAELARPGASVKMSCKASGYTFTRYTMHWVKQRPGQGLEWIGYINPSRGY TNYNQKFKDKATLTTDKSSSTAYMELTRLTSEDSAVYYCARYYDDHYCLDYWGQGTT VTVSS (SEQ ID NO: 266).

The light chain variable region ($V_L$) of the Hit3a antibody, with CDRs underlined, is:

DIELTQSPAIMSASPGEKVTMTCSASSSVSYMNWYQQKSGTSPKRWIYDTSKLASGVPA RFSGSGSGTSYSLTISGMEAEDAATYYCQQWSSNPFTFGSGTKLELK (SEQ ID NO: 267).

**[0209]** In some embodiments, the anti-CD3 scFv or anti-CD3 Fab comprises the heavy chain and light chain CDRs in a heterologous heavy chain and light chain framework region, respectively, or the entire heavy and light chain variable regions, of the 12F6 antibody (see, e.g., Li et al., *Immunology* 116: 4 (2005) 487-498). The heavy chain variable region ($V_H$) of the 12F6 antibody, with CDRs underlined, is:

QVQLQQSGAELARPGASVKMSCKASGYTFTSYTMHWVKQRPGQGLEWIGYINPSSGYT KYNQKFKDKATLTADKSSSTAYMQLSSLTSEDSAVYYCARWQDYDVYFDYWGQGTTL TVSS (SEQ ID NO: 268).

The light chain variable region ($V_L$) of the 12F6 antibody, with CDRs underlined, is:

QIVLSQSPAILSASPGEKVTMTCRASSSVSYMHWYQQKPGSSPKPWIYATSNLASGVPA RFSGSGSGTSYSLTISRVEAEDAATYYCQQWSSNPPTFGGGTKLETK (SEQ ID NO: 269).

**[0210]** In some embodiments, the anti-CD3 scFv or anti-CD3 Fab comprises the heavy chain and light chain CDRs in a heterologous heavy chain and light chain framework region, respectively, or the entire heavy and light chain variable regions, of the SP34 antibody (see, e.g., U.S. Patent No. 12,006,367). The heavy chain variable region ($V_H$) of the SP34 antibody, with CDRs underlined, is:

EVQLVESGGGLVQPKGSLKLSCAASGFTFNTYAMNWVRQAPGKGLEWVARIRSKYNN YATYYADSVKDRFTISRDDSQSILYLQMNNLKTEDTAMYYCVRHGNFGNSYVSWFAY WGQGTLVTVSA (SEQ ID NO: 270).

The light chain variable region ($V_L$) of the SP34 antibody, with CDRs underlined, is:

QAVVTQESALTTSPGETVTLTCRSSTGAVTTSNYANWVQEKPDHLFTGLIGGTNKRAPG VPARFSGSLIGDKAALTITGAQTEDEAIYFCALWYSNLWVFGGGTKLTVL (SEQ ID NO: 271).

**[0211]** In some embodiments, the anti-CD3 scFv or anti-CD3 Fab comprises the heavy chain and light chain CDRs in a heterologous heavy chain and light chain framework region, respectively, or the entire heavy and light chain variable regions, of the UCHT1 antibody (see, e.g., Arnett, et al., *PNAS USA* 101 (46) 16268-16273 (2004)). The heavy chain variable region ($V_H$) of the UCHT1 antibody, with CDRs underlined, is:

EVQLQQSGPELVKPGASMKISCKASGYSFTGYTMNWVKQSHGKNLEWMGLINPYKGV STYNQKFKDKATLTVDKSSSTAYMELLSLTSEDSAVYYCARSGYYGDSDWYFDVWGQ GTTLTVFS (SEQ ID NO: 272).

The light chain variable region ($V_L$) of the UCHT1 antibody, with CDRs underlined, is:

DIQMTQTTSSLSASLGDRVTISCRASQDIRNYLNWYQQKPDGTVKLLIYYTSRLHSGVPS KFSGSGSGTDYSLTISNLEQEDIATYFCQQGNTLPWTFAGGTKLEIK (SEQ ID NO: 273).

**[0212]** In some embodiments, the anti-CD3 scFv or anti-CD3 Fab comprises the heavy chain and light chain CDRs in a heterologous heavy chain and light chain framework region, respectively, or the entire heavy and light chain variable regions, of the Acapatamab antibody (see, e.g., Dorff, Tanya et al. Clinical cancer research vol. 30,8 (2024): 1488-1500. doi:10.1158/1078-0432.CCR-23-2978). The heavy chain variable region ($V_H$) of the Acapatamab antibody, with CDRs underlined, is:

EVQLVESGGGLVQPGGSLKLSCAASGFTFNKYAMNWVRQAPGKGLEWVARIRSKYNN YATYYADSVKDRFTISRDDSKNTAYLQMNNLKTEDTAVYYCVRHGNFGNSYISYWAY WGQGTLVTVSS (SEQ ID NO: 279).

The light chain variable region ($V_L$) of the Acapatamab antibody, with CDRs underlined, is:

QTVVTQEPSLTVSPGGTVTLTCGSSTGAVTSGNYPNWVQQKPGQAPRGLIGGTKFLAPG TPARFSGSLLGGKAALTLSGVQPEDEAEYYCVLWYSNRWVFGGGTKLTVL (SEQ ID NO: 280).

**[0213]** In some embodiments, the anti-CD3 scFv or anti-CD3 Fab comprises the heavy chain and light chain CDRs in a heterologous heavy chain and light chain framework region, respectively, or the entire heavy and light chain variable regions, of the hOKT3/teplizumab antibody (see, e.g., TZield® (Sanofi); Masharani UB, Becker J. Expert Opin Biol Ther. 2010;10(3):459-465. doi:10.1517/14712591003598843). The heavy chain variable region ($V_H$) of the hOKT3/teplizumab antibody, with CDRs underlined, is:

QVQLVQSGGGVVQPGRSLRLSCKASGYTFTRYTMHWVRQAPGKGLEWIGYINPSRGYT NYNQKVKDRFTISRDNSKNTAFLQMDSLRPEDTGVYFCARYYDDHYCLDYWGQGTPV TVSS (SEQ ID NO: 281).

The light chain variable region ($V_L$) of the hOKT3/teplizumab antibody, with CDRs underlined, is:

DIQMTQSPSSLSASVGDRVTITCSASSSVSYMNWYQQTPGKAPKRWIYDTSKLASGVPS RFSGSGSGTDYTFTISSLQPEDIATYYCQQWSSNPFTFGQGTKLQIT (SEQ ID NO: 282).

**[0214]** In some embodiments, the anti-CD3 scFv or anti-CD3 Fab comprises the heavy chain and light chain CDRs in a heterologous heavy chain and light chain framework region, respectively, or the entire heavy and light chain variable regions, of the FN-18 antibody (see, e.g., Sen KI, Tang WH, Nayak S, et al. J Am Soc Mass Spectrom. 2017;28(5):803-810. doi:10.1007/s13361-016-1580-0). The heavy chain variable region ($V_H$) of the FN-18 antibody, with CDRs underlined, is:

QVQLQQSEAELARPGASVKMSCKASGYTFTDYTIHWLKQRPGQGLDWIGYFNPSSESTE YNRKFKDRTILTADRSSTTAYMQLSSLTSEDSAVYYCSRKGEKLLGNRYWYFDVWGAG TSVTVSS (SEQ ID NO: 283).

The light chain variable region ($V_L$) of the FN-18 antibody, with CDRs underlined, is:

DIVMSQSPSSLAVSVGEKVTMSCKSSQSLLYSSNQKNYLAWYQQKPGQSPKLLINWAST RESGVPDRFTGSGSRTDFTLTISSVKAEDLAVYFCQQFYSYPPTFGGGTKLEIK (SEQ ID NO: 284).

**[0215]** In some embodiments, the anti-CD3 scFv or anti-CD3 Fab comprises the heavy chain and light chain CDRs in a heterologous heavy chain and light chain framework region, respectively, or the entire heavy and light chain variable regions, of the SP34-2 antibody (see, e.g., Sen KI, Tang WH, Nayak S, et al. J Am Soc Mass Spectrom. 2017;28(5):803-810. doi:10.1007/s13361-016-1580-0). The heavy chain variable region ($V_H$) of the SP34-2 antibody, with CDRs underlined, is:

EVKLLESGGGLVQPKGSLKLSCAASGFTFNTYAMNWVRQAPGKGLEWVARIRSKYNN YATYYADSVKDRFTISRDDSQSILYLQMNNLKTEDTAMYYCVRHGNFGNSYVSWFAY WGQGTLVTVSA (SEQ ID NO: 285).

The light chain variable region ($V_L$) of the SP34-2 antibody, with CDRs underlined, is:

QAVVTQESALTTSPGETVTLTCRSSTGAVTTSNYANWVQEKPDHLFTGLIGGTNKRAPG VPARFSGSLIGDKAALTITGAQTEDEAIYFCALWYSNLWVFGGGTKLTVL (SEQ ID NO: 286).

**[0216]** In some embodiments, the targeting polypeptide comprises an anti-CD28 antibody or fragment thereof, e.g., an anti-CD28 scFv or anti-CD28 Fab. In some embodiments, the anti-CD3 scFv or anti-CD3 Fab comprises the heavy chain and light chain CDRs in a heterologous heavy chain and light chain framework region, respectively, or the entire heavy and light chain variable regions, of the CD28 (SA) antibody. The heavy chain variable region ($V_H$) of the CD28 (SA) antibody, with CDRs underlined, is:

QVQLVQSGAEVKKPGASVKVSCKASGYTFTSYYIHWVRQAPGQGLEWIGCIYPGNVNT NYNEKFKDRATLTVDTSISTAYMELSRRSDDTAVYFCTRSHYGLDWNFDVWGQGTTVT VSS (SEQ ID NO: 274).

The light chain variable region ($V_H$) of the CD28 (SA) antibody, with CDRs underlined, is:

DIQMTQSPSSISASVGDRVTITCHASQNIYVWLNWYQQKPGKAPKLLIYKASNLHTGVP SRESGSGSGTDFTLTISSLQPEDFATYYCQQGQTYPYTFGGGTKVEIK (SEQ ID NO: 275).

**[0217]** In some embodiments, the targeting polypeptide is a CD28 natural ligand. In some embodiments, the CD28 natural ligand is CD80 or CD86 or a CD28-binding fragment thereof. The CD80 amino acid sequence is:

MGHTRRQGTSPSKCPYLNFFQLLVLAGLSHFCSGVIHVTKEVKEVATLSCGHNVSVEEL AQTRIYWQKEKKMVLTMMSGDMNIWPEYKNRTIFDITNNLSIVILALRPSDEGTYECVV LKYEKDAFKREHLAEVTLSVKADFPTPSISDFEIPTSNIRRIICSTSGGFPEPHLSWLENGE ELNAINTTVSQDPETELYAVSSKLDFNMTTNHSFMCLIKYGHLRVNQTFNWNTTKQEHF PDNLLPSWAITLISVNGIFVICCLTYCFAPRCRERRRNERLRRESVRPV (SEQ ID NO: 276).

The CD86 amino acid sequence is:

MDPQCTMGLSNILFVMAFLLSGAAPLKIQAYFNETADLPCQFANSQNQSLSELVVFWQD QENLVLNEVYLGKEKFDSVHSKYMGRTSFDSDSWTLRLHNLQIKDKGLYQCIIHHKKPT GMIRIHQMNSELSVLANFSQPEIVPISNITENVYINLTCSSIHGYPEPKKMSVLLRTKNSTI EYDGVMQKSQDNVTELYDVSISLSVSFPDVTSNMTIFCILETDKTRLLSSPFSIELEDPQPP PDHIPWITAVLPTVIICVMVFCLILWKWKKKRPRNSYKCGTNTMEREESEQTKKREKIH IPERSDEAQRVFKSSKTSSCDKSDTCF (SEQ ID NO: 277).

**[0218]** In some embodiments, the targeting polypeptide is a 4-1BB antibody or natural ligand. In some embodiments, the targeting polypeptide comprises an anti-4-1BB antibody or fragment thereof, e.g., an anti-4-1BB scFv. In some embodiments, the 4-1BB natural ligand is CD137 (41BBL) or a 4-1BB-binding fragment thereof. The amino acid sequence of CD137 (41BBL) is

MEYASDASLDPEAPWPPAPRARACRVLPWALVAGLLLLLLLAAACAVFLACPWAVSG ARASPGSAASPRLREGPELSPDDPAGLLDLRQGMFAQLVAQNVLLIDGPLSWYSDPGLA GVSLTGGLSYKEDTKELVVAKAGVYYVFFQLELRRVVAGEGSGSVSLALHLQPLRSAA GAAALALTVDLPPASSEARNSAFGFQGRLLHLSAGQRLGVHLHTEARARHAWQLTQGA TVLGLFRVTPEIPAGLPSPRSE (SEQ ID NO: 278).

**[0219]** In some cases where the CDRs of an antibody or fragment thereof for use in a targeting polypeptide is not disclosed herein, the CDRs may be determined using the Kabat numbering system. *See, e.g.,* system Kabat et al. (1991) Sequences of Proteins of Immunological Interest Fifth Edition, U.S. Department of Health and Human Services, NIH Publication No. 91-3242.

**[0220]** In some embodiments, the EDV selectively targets B cells. In some embodiments, the targeting polypeptide binds to CD19, CD20, BCMA, CD138, TACI, or CD22.

**[0221]** In some embodiments, the EDV selectively targets NK cells. In some embodiments, the targeting polypeptide binds to CD56, CD16, NKp46/NCR1, NCR2, or KIR.

**[0222]** In some embodiments, the EDV selectively targets monocytes or macrophages. In some embodiments, the targeting polypeptide binds to CD11b, CD68, CD14, CD33, or CD163.

**[0223]** In some embodiments, the EDV selectively targets dendritic cells. In some embodiments, the targeting polypeptide binds to CD11b, CD11c, XCR1, CD33, CD1c, or CD123.

**[0224]** In some embodiments, the EDV selectively targets HSCs. In some embodiments, the targeting polypeptide binds to CD34, CD117, CD49f, CD38, CD90, or EPCR.

**[0225]** In some embodiments, the targeting polypeptide is a fusion polypeptide comprising: (i) a target-binding region, e.g., an antibody or fragment thereof; and (ii) a heterologous polypeptide (a "fusion partner"). The fusion partner can be a polypeptide that enhances accessibility of the antibody to a target cell. Suitable fusion partners include, but are not limited to, the stalk portion of a polypeptide; the stalk and transmembrane domain of a polypeptide; an immunoglobulin hinge polypeptide; a linker polypeptide; and the like. In some embodiments, the targeting polypeptide is fused to a transmembrane domain via a linker.

**[0226]** In some cases, the fusion partner is the stalk and transmembrane domain of a transmembrane protein. A "transmembrane domain" (TMD), as used herein, is a portion of a transmembrane (TM) protein that contains a hydrophobic portion that can insert into or span a cell membrane. Transmembrane components or domains have a three-dimensional structure that is thermodynamically stable in a cell membrane and generally range in length from about 15 amino acids to about 30 amino acids. The structure of a transmembrane component or domain may comprise an alpha helix, a beta barrel, a beta sheet, a beta helix, or any combination thereof. In

certain embodiments, a transmembrane component or domain comprises or is derived from a known transmembrane protein, e.g., a CD4 transmembrane domain, a CD8 transmembrane domain, a CD27 transmembrane domain, a CD28 transmembrane domain, or any combination thereof.

**[0227]** In some embodiments, the fusion partner includes the stalk and transmembrane domain of a CD8α polypeptide. In some embodiments, the stalk and transmembrane domain can comprise the amino acid sequence as set forth in SEQ ID NO: 92 (shown below) where the stalk has the amino acid sequence as set forth in SEQ ID NO: 93 (shown below) and the TMD has the amino acid sequence as set forth in SEQ ID NO: 94 (shown below).

TTTPAPRPPTPAPTIASQPLSLRPEACRPAAGGAVHTRGLDFACDIYIWAPLAGTCGVLLL SLVITLYC (SEQ ID NO: 92)

TTTPAPRPPTPAPTIASQPLSLRPEACRPAAGGAVHTRGLDFACD (SEQ ID NO: 93)

IYIWAPLAGTCGVLLLSLVITLYC (SEQ ID NO: 94)

**[0228]** In some embodiments, the fusion partner includes the stalk domain of a CD8α polypeptide comprising the amino acid sequence as set forth in SEQ ID NO: 93 (shown above). In some embodiments, the fusion partner includes the stalk domain of a CD8α polypeptide comprising the amino acid sequence as set forth in SEQ ID NO: 95 (shown below).

ASAKPTTTPAPRPPTPAPTIASQPLSLRPEAARPAAGGAVHTRGLDFAK (SEQ ID NO: 95)

**[0229]** In some embodiments, the fusion partner comprises a polypeptide linker. Many polypeptide linkers may be used and non-limiting examples are discussed in detail below. In some embodiments, the polypeptide linker comprises a glycine-rich polypeptide having a length of from 5 amino acids to about 50 amino acids; e.g., where the fusion partner comprises the sequence (GGGGS)n (SEQ ID NO: 216), where n is an integer from 1 to 10, e.g., where n is 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10, e.g., where n is 3. Polypeptide linkers are discussed in detail below.

**[0230]** In some embodiments, the fusion partner includes an immunoglobulin (Ig) hinge polypeptide. As used herein, a "hinge polypeptide," "hinge region," or a "hinge" refers to (a) an immunoglobulin hinge sequence (made up of, for example, upper and core regions of an immunoglobulin hinge) or a functional fragment or variant thereof, (b) a type II C-lectin interdomain (stalk) region or a functional fragment or variant thereof, or (c) a cluster of

differentiation (CD) molecule stalk region or a functional variant thereof. As used herein, a "wild-type immunoglobulin hinge region" refers to a naturally occurring upper and middle hinge amino acid sequences interposed between and connecting the CH1 and CH2 domains (for IgG, IgA, and IgD) or interposed between and connecting the CH1 and CH3 domains (for IgE and IgM) found in the heavy chain of an antibody.

### iv. Exemplary EDVs

**[0231]** In some embodiments, the EDV comprises a gag protein fragment that comprises an MA polypeptide and a CA polypeptide. In some embodiments the gag protein fragment comprises MA polypeptide with the sequence as set forth in (SEQ ID NO: 59). In some embodiments the gag protein fragment comprises CA polypeptide with an amino acid sequence that corresponds to amino acids at positions 149-231 of a (SEQ ID NO: 63).

**[0232]** In some embodiments, the EDV gag protein fragment comprises an MA polypeptide, a CA polypeptide, a p2 polypeptide, an NC polypeptide, a p1 polypeptide, and a p6 polypeptide. In some embodiments, the EDV gag protein fragment comprises an MA polypeptide, a CA polypeptide, a p2 polypeptide, an amino acid sequence IQKGRQAN (SEQ ID NO: 73), a p1 polypeptide, and a p6 polypeptide. In some embodiments, the gag protein fragment (e.g., derived from an HIV gag protein) includes one or more heterologous protease cleavage sites between one or more of: i) the MA polypeptide and the CA polypeptide; ii) the CA polypeptide and the p2 polypeptide; iii) the p2 polypeptide and the NC polypeptide; iv) the IQKGRQAN (SEQ ID NO: 73) and the p1 polypeptide; and v) the p1 polypeptide and the p6 polypeptide. Protease cleavage sites and polypeptide sequences are discussed below.

**[0233]** In some embodiments, the EDV comprises a VSVG viral envelope protein with the sequence as set forth in SEQ ID NO: 80 and a polynucleotide-guided nuclease. In some embodiments, the polynucleotide-guided nuclease is CRISPR-Cas9.

**[0234]** In some embodiments, the EDV comprises a viral envelope protein VSVGmut with the sequence as set forth in SEQ ID NO: 81 and a polynucleotide-guided nuclease. In some embodiments, the polynucleotide-guided nuclease is CRISPR-Cas9.

**[0235]** In some embodiments, the EDV further comprises a guide polynucleotide. In some embodiments, the guide polynucleotide is a gRNA. In these embodiments, the CRISPR-Cas9 and the gRNA form an RNP.

**[0236]** In some embodiments, the gRNA is targeted to the *TRAC* locus. In some embodiments, the gRNA is targeted to exon 1 of the *TRAC* locus. In some embodiments, the gRNA comprises the sequence of CAGGGTTCTGGATATCTGT (SEQ ID NO: 137) or TCAGGGTTCTGGATATCTGT (SEQ ID NO: 138). In some embodiments, the gRNA is under control of the U6 promoter. In some embodiments, the EDV vector comprises a U6 promoter-TRAC-sgRNA scaffold comprises the sequence as set forth in SEQ ID NO: 139.

GAGGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGATACAAGGCTGTTAGA
GAGATAATTAGAATTAATTTGACTGTAAACACAAAGATATTAGTACAAAATACGTG
ACGTAGAAAGTAATAATTTCTTGGGTAGTTTGCAGTTTTAAAATTATGTTTTAAAAT
GGACTATCATATGCTTACCGTAACTTGAAAGTATTTCGATTTCTTGGCTTTATATATC
TTGTGGAAAGGACGAAACACCGCAGGGTTCTGGATATCTGTGTTTTAGAGCTAGAA
ATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTC
GGTGCTTTTTT (SEQ ID NO: 139).

**[0237]** In some embodiments, the EDV further comprises a targeting polypeptide. In some embodiments, the targeting polypeptide binds to CD3. In some embodiments, the targeting polypeptide comprises an anti-CD3 antibody or fragment thereof, e.g., an anti-CD3 scFv or anti-CD3 Fab . In some embodiments, the anti-CD3 scFv or anti-CD3 Fab comprises the amino acid sequence as set forth in SEQ ID NO: 97 as shown below. In some embodiments, the targeting polypeptide, e.g., anti-CD3 scFv or anti-CD3 Fab, is covalently linked to a CD8α stalk polypeptide, a CD8α stalk and transmembrane domain, a PDFGR transmembrane domain, or a PDFGR transmembrane domain and a CD8α stalk polypeptide. In some embodiments, the targeting polypeptide further comprises a linker polypeptide sequence.

MALPVTALLLPLALLLHAARPQVQLQQSGAELARPGASVKMSCKASGYTFTRYTMHW
VKQRPGQGLEWIGYINPSRGYTNYNQKFKDKATLTTDKSSSTAYMQLSSLTSEDSAVYY
CARYYDDHYCLDYWGQGTTLTVSSSGGGGSGGGGSGGGGSIVLTQSPAIMSASPGEKV
TMTCSASSSVSYMNWYQQKSGTSPKRWIYDTSKLASGVPAHFRGSGSGTSYSLTISGME
AEDAATYYCQQWSSNPFTFGSGTKLEINRAA (SEQ ID NO: 97)

### c. Lipid Nanoparticles (LNPs)

**[0238]** Disclosed herein are lipid nanoparticles (LNPs) for use in conjunction with a virus vector (e.g., an AAV vector) of the present disclosure. The LNP can include or be composed of lipid

particles and/or liposomes. LNPs may encapsulate nucleic acids within cationic lipid particles (e.g., liposomes) for delivery to target cells. Non-limiting examples of ionizable lipids that may be used in an LNP include SM-102 (CAS number: 2089251-47-6), Dlin-MC3-DMAMC3 (also referred to as "MC3;" CAS number: 1224606-06-7), CL4H6 (CAS number: 2256087-35-9), ssPalm-O-Phe (CAS number: 2377474-67-2), ALC-0315 (CAS number: 2036272-55-4), and LP01 (CAS number: 1799316-64-5). Non-limiting examples of neutral lipids that may be used in an LNP include 18:1 Δ9-cis phosphoethanolamine (DOPE) (CAS number: 4004-05-1), cholesterol (CAS number: 57-88-5), and 14:0 PEG2000 phosphoethanolamine (C14-PEG2000; CAS number: 474922-82-2). In some embodiments, LNPs contain viral particles and/or viral components to increase transduction efficiency. In some embodiments, LNPs do not contain any viral components, which helps minimize safety and immunogenicity concerns. Lipid particles may be used for *in vitro, ex vivo,* and *in vivo* deliveries. In some embodiments, an LNP can be delivered to a subject *in vivo*. See, for example, U.S. Patent No. 9,737,604 and Zhang et al. "Lipid nanoparticle-mediated efficient delivery of CRISPR/Cas9 for tumor therapy," NPG Asia Materials Volume 9, page e441 (2017).

**[0239]** In some embodiments, the LNP comprises a polynucleotide-guided nuclease, or polynucleotide encoding the nuclease, for delivery to a target cell. Many suitable gene-editing nucleases may be used as a polynucleotide-guided nuclease and are discussed in detail below. Suitable nucleases include, but are not limited to, a homing nuclease polypeptide; a FokI polypeptide; a transcription activator-like effector nuclease (TALEN) polypeptide; a MegaTAL polypeptide; a meganuclease polypeptide; a zinc finger nuclease (ZFN); an ARCUS nuclease; and the like. The meganuclease can be engineered from an LADLIDADG homing endonuclease (LHE). A megaTAL polypeptide can comprise a TALE DNA binding domain and an engineered meganuclease. See, e.g., WO 2004/067736 (homing endonuclease); Urnov et al. (2005) Nature 435:646 (ZFN); Mussolino et al. (2011) Nucle. Acids Res. 39:9283 (TALE nuclease); Boissel et al. (2013) Nucl. Acids Res. 42:2591 (MegaTAL). In some embodiments, the polynucleotide-guided nuclease is a CRISPR-Cas nuclease. In some embodiments, the polynucleotide-guided nuclease is CRISPR-Cas9, CAST, CRISPR-directed integrase, CRISPR-directed recombinase, or variant thereof. In some embodiments the LNP comprises a DNA, or RNA encoding the nuclease.

**[0240]** In some embodiments, the LNP also comprises a targeting polypeptide, optionally fused to a transmembrane domain that anchors the targeting polypeptide in the LNP, enabling binding of the LNP to a target cell.

**[0241]** In some embodiments, the LNP also comprises a guide polynucleotide. In some embodiments, the guide polynucleotide forms a complex with the polynucleotide-guided nuclease.

**[0242]** In some cases, LNPs may also be used to deliver a polynucleotide that comprises a coding sequence to a target cell. For example, an LNP may comprise a coding sequence for the gag protein fragment, a viral envelope protein, a polynucleotide-guided nuclease, and/or a targeting polypeptide. In some cases, the LNP may comprise a coding sequence for a guide polynucleotide.

### i. Targeting Polypeptide

**[0243]** Targeting polypeptides of an LNP enable delivery of genome-editing molecules to specific target cells. Any targeting polypeptide as described above in the context of EDV can be used in an LNP.

### ii. Lipids and Liposomes

**[0244]** Components in LNPs may comprise cationic lipids 1,2- dilineoyl-3-dimethylammonium-propane (DLinDAP), l,2-dilinoleyloxy-3-N,N- dimethylaminopropane (DLinDMA), l,2-dilinoleyloxyketo-N,N-dimethyl-3-aminopropane (DLinK-DMA), l,2-dilinoleyl-4-(2-dimethylaminoethyl)-[l,3]-dioxolane (DlinKC2-DMA), (3- o-[2"-(methoxypolyethyleneglycol 2000) succinoyl]-l,2-dimyristoyl-sn-glycol (PEG-S-DMG), R-3-[(ro-methoxy-poly(ethylene glycol)2000) carbamoyl]-l,2-dimyristyloxlpropyl-3-amine (PEGC-DOMG, and any combination thereof. Preparation of LNPs and encapsulation may be adapted from Rosin et al, Molecular Therapy, vol. 19, no. 12, pages 1286-2200, Dec. 2011). The LNPs may also comprise one or more other types of lipids, e.g., cationic lipids, such as amino lipid 2,2-dilinoleyl-4-dimethylaminoethyl-[l,3]- dioxolane (Dlin-KC2-DMA), Dlin-KC2-DMA4, C12- 200 and colipids disteroylphosphatidyl choline, cholesterol, and PEG-DMG. In some embodiments, the LNP is a lipidoid, such as any of those set forth in, for example, U.S. Patent Application Publication No. 2011/0293703. In some embodiments, the LNP comprises an amino lipid, such as any of those set forth in, for example, Jayaraman, Angew. Chem. Int. Ed. 2012, 51, 8529 –8533. In some

embodiments, the LNP comprises a lipid envelope, such as any of those set forth in, for example, Korman et al., 2011. Nat. Biotech. 29:154-157.

**[0245]** In some embodiments, the LNP contains a nucleic acid, wherein the charge ratio of nucleic acid backbone phosphates to cationic lipid nitrogen atoms is about 1: 1.5 – 7 or about 1:4.

**[0246]** In some embodiments, the LNP also includes a shielding compound, which is removable from the lipid composition under *in vivo* conditions. In some embodiments, the shielding compound is a biologically inert compound. In some embodiments, the shielding compound does not carry any charge on its surface or on the molecule as such. In some embodiments, the shielding compounds are polyethylenglycols (PEGs), hydroxyethylglucose (HEG) based polymers, polyhydroxyethyl starch (polyHES) and polypropylene. In some embodiments, the PEG, HEG, polyHES, and a polypropylene weight between about 500 to 10,000 Da or between about 2000 to 5000 Da. In some embodiments, the shielding compound is PEG 2000 or PEG 5000.

**[0247]** In some embodiments, the LNP can include one or more helper lipids. In some embodiments, the helper lipid can be a phosphor lipid or a steroid. In some embodiments, the helper lipid is between about 20 mol % to 80 mol % of the total lipid content of the composition. In some embodiments, the helper lipid component is between about 35 mol % to 65 mol % of the total lipid content of the LNP. In some embodiments, the LNP includes lipids at 50 mol% and the helper lipid at 50 mol% of the total lipid content of the LNP.

**[0248]** In some embodiments, a lipid particle may be liposome. Liposomes are spherical vesicle structures composed of a unilamellar or multilamellar lipid bilayer surrounding internal aqueous compartments and a relatively impermeable outer lipophilic phospholipid bilayer. In some embodiments, liposomes are biocompatible, nontoxic, can deliver both hydrophilic and lipophilic drug molecules, protect their cargo from degradation by plasma enzymes, and transport their load across biological membranes and the blood brain barrier (BBB).

**[0249]** Liposomes can be made from several different types of lipids, e.g., phospholipids. A liposome may comprise natural phospholipids and lipids such as 1,2-distearoryl-sn-glycero-3-phosphatidyl choline (DSPC), sphingomyelin, egg phosphatidylcholines, monosialoganglioside, or any combination thereof.

**[0250]** Several other additives may be added to liposomes in order to modify their structure and properties. For instance, liposomes may further comprise cholesterol, sphingomyelin, and/or l,2-dioleoyl-sn-glycero-3-phosphoethanolamine (DOPE), e.g., to increase stability and/or to prevent the leakage of the liposomal inner cargo.

**[0251]** Other non-limiting, exemplary liposomes can be those as set forth in Wang et al., ACS Synthetic Biology, 1, 403-07 (2012); Wang et al., PNAS, 113(11) 2868-2873 (2016); Spuch and Navarro, Journal of Drug Delivery, vol. 2011, Article ID 469679, 12 pages, 2011. Doi:10.1155/2011/469679; WO 2008/042973; US Pat. No. 8,071,082; WO 2014/186366; 20160257951; US 20160129120; US 20160244761; US 20120251618; WO 2013/093648; Lipofectin (a combination of DOTMA and DOPE), Lipofectase, LIPOFECTAMINE.RTM. (e.g., LIPOFECTAMINE.RTM. 2000, LIPOFECTAMINE.RTM. 3000, LIPOFECTAMINE.RTM. RNAiMAX, LIPOFECTAMINE.RTM. LTX), SAINT-RED (Synvolux Therapeutics, Groningen Netherlands), DOPE, Cytofectin (Gilead Sciences, Foster City, Calif.), and Eufectins (JBL, San Luis Obispo, Calif.).

## III. Gene Editing Systems

### a. Polynucleotide-Guided Nucleases

**[0252]** Disclosed herein are polynucleotide-guided nucleases for modifying the genome of a target cell *in vivo* as delivered by an EDV or LNP as discussed above.

**[0253]** In many embodiments, the polynucleotide-guided nuclease is nuclease or variant thereof from a gene editing system. Suitable polynucleotide-guided nucleases include, but are not limited to, a homing nuclease polypeptide; a FokI polypeptide; a transcription activator-like effector nuclease (TALEN) polypeptide; a MegaTAL polypeptide; a meganuclease polypeptide; a zinc finger nuclease (ZFN); an ARCUS nuclease; and the like. The meganuclease can be engineered from an LADLIDADG homing endonuclease (LHE). A megaTAL polypeptide can comprise a TALE DNA binding domain and an engineered meganuclease. See, e.g., WO 2004/067736 (homing endonuclease); Urnov et al. (2005) Nature 435:646 (ZFN); Mussolino et al. (2011) Nucle. Acids Res. 39:9283 (TALE nuclease); Boissel et al. (2013) Nucl. Acids Res. 42:2591 (MegaTAL).

**[0254]** In some embodiments, the polynucleotide-guided nuclease is a CRISPR-Cas nuclease. In general, the CRISPR-Cas nuclease performs dsDNA cleavage, a single nick, or a dual nick at a

target site. In some embodiments, the CRISPR-Cas nuclease forms a ribonucleoprotein (RNP) complex with a guide polynucleotide.

**[0255]** The CRISPR-Cas nuclease can be any of a variety of CRISPR-Cas nucleases. CRISPR-Cas nucleases can be derived from a variety of bacterial species including, but not limited to, *Veillonella atypical*, *Fusobacterium nucleatum*, *Filifactor alocis*, *Solobacterium moorei*, *Coprococcus catus*, *Treponema denticola*, *Peptoniphilus duerdenii*, *Catenibacterium mitsuokai*, *Streptococcus mutans*, *Listeria innocua*, *Staphylococcus pseudintermedius*, *Acidaminococcus intestine*, *Olsenella uli*, *Oenococcus kitaharae*, *Bifidobacterium bifidum*, *Lactobacillus rhamnosus*, *Lactobacillus gasseri*, *Finegoldia magna*, *Mycoplasma mobile*, *Mycoplasma gallisepticum*, *Mycoplasma ovipneumoniae*, *Mycoplasma canis*, *Mycoplasma synoviae*, *Eubacterium rectale*, *Streptococcus thermophilus*, *Eubacterium dolichum*, *Lactobacillus coryniformis* subsp. *Torquens*, *Ilyobacter polytropus*, *Ruminococcus albus*, *Akkermansia muciniphila*, *Acidothermus cellulolyticus*, *Bifidobacterium longum*, *Bifidobacterium dentium*, *Corynebacterium diphtheria*, *Elusimicrobium minutum*, *Nitratifractor salsuginis*, *Sphaerochaeta globus*, *Fibrobacter succinogenes subsp. Succinogenes*, *Bacteroides fragilis*, *Capnocytophaga ochracea*, *Rhodopseudomonas palustris*, *Prevotella micans*, *Prevotella ruminicola*, *Flavobacterium columnare*, *Aminomonas paucivorans*, *Rhodospirillum rubrum*, *Candidatus Puniceispirillum marinum*, *Verminephrobacter eiseniae*, *Ralstonia syzygii*, *Dinoroseobacter shibae*, *Azospirillum*, *Nitrobacter hamburgensis*, *Bradyrhizobium*, *Wolinella succinogenes*, *Campylobacter jejuni* subsp. *Jejuni*, *Helicobacter mustelae*, *Bacillus cereus*, *Acidovorax ebreus*, *Clostridium perfringens*, *Parvibaculum lavamentivorans*, *Roseburia intestinalis*, *Neisseria meningitidis*, *Pasteurella multocida subsp. Multocida*, *Sutterella wadsworthensis*, *proteobacterium*, *Legionella pneumophila*, *Parasutterella excrementihominis*, *Wolinella succinogenes*, and *Francisella novicida*. Suitable CRISPR-Cas nucleases are described in detail below. Examples of CRISPR-Cas nucleases are CRISPR-Cas endonucleases (e.g., class 2 CRISPR-Cas nucleases such as a type II, type V, or type VI CRISPR-Cas nuclease). In some embodiments, the CRISPR-Cas nuclease is a type II CRISPR-Cas nuclease. In some embodiments, the type II CRISPR-Cas nuclease is a Cas9 polypeptide. In some embodiments, the CRISPR-Cas nuclease is a type V CRISPR-Cas nuclease, e.g., a Cas12a, a Cas12b, a Cas12c, a Cas12d, a Cas12e, a Cpf1, a C2c1, or a C2c3 polypeptide. In some embodiments, the CRISPR-Cas nuclease is a type VI CRISPR-Cas nuclease, e.g., a Cas13a, a Cas13b, a Cas13c, a Cas13d, a C2c2 (also

referred to as Cas13a) polypeptide. In some embodiments, the CRISPR-Cas nuclease is a Cas14 polypeptide. In some embodiments, the CRISPR-Cas nuclease is a Cas14a polypeptide, a Cas14b polypeptide, or a Cas14c polypeptide. In some embodiments, a suitable CRISPR-Cas nuclease is a CasX or a CasY polypeptide. CasX and CasY polypeptides are described in Burstein et al. (2017) Nature 542:237.

**[0256]** Also suitable for use is a variant CRISPR-Cas nuclease, where the variant is a high-fidelity or enhanced specificity CRISPR-Cas nuclease with reduced off-target effects and robust on-target cleavage. Non-limiting examples of CRISPR-Cas nuclease variants with improved on-target specificity include the SpCas9 (K855A), SpCas9 (K810A/K1003A/R1060A) (also referred to as eSpCas9(1.0)), and SpCas9 (K848A/K1003A/R1060A) (also referred to as eSpCas9(1.1)) variants described in Slaymaker *et al.*, *Science*, 351(6268):84-8 (2016), and the SpCas9 variants described in Kleinstiver *et al.*, *Nature*, 529(7587):490-5 (2016) containing one, two, three, or four of the following mutations: N497A, R661A, Q695A, and Q926A (*e.g.*, SpCas9-HF1 contains all four mutations).

**[0257]** Also suitable for use, e.g., when fused with a second enzyme with nicking of DNA cleaving activity, is a variant CRISPR-Cas effector polypeptide, where the variant CRISPR-Cas effector polypeptide has reduced or no nucleic acid cleavage activity. For example, the dCas9 variant (Jinek et al., Science, 2012, 337:816-821; Qi et al., Cell, 152(5):1173-1183) contains two silencing mutations of the RuvC1 and HNH nuclease domains (D10A and H840A). In some embodiments, the dCas9 polypeptide from Streptococcus pyogenes comprises at least one mutation at position D10, G12, G17, E762, H840, N854, N863, H982, H983, A984, D986, A987 or any combination thereof. Descriptions of such dCas9 polypeptides and variants thereof are provided in, for example, International Patent Publication No. WO 2013/176772. The dCas9 enzyme can contain a mutation at D10, E762, H983, or D986, as well as a mutation at H840 or N863. In some instances, the dCas9 enzyme can contain a D10A or D10N mutation. Also, the dCas9 enzyme can contain a H840A, H840Y, or H840N. In some embodiments, the dCas9 enzyme can contain D10A and H840A; D10A and H840Y; D10A and H840N; D10N and H840A; D10N and H840Y; or D10N and H840N substitutions. The substitutions can be conservative or non-conservative substitutions to render the Cas9 polypeptide catalytically inactive and able to bind to target nucleic acid.

**[0258]** In some embodiments, a suitable CRISPR-Cas nuclease comprises an amino acid sequence having at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, or 100%, amino acid sequence identity to the amino acid sequence as set forth in any one of SEQ ID NOs: 98- as shown in Table 3 below.

**Table 3 – CRISPR-Cas Nuclease Sequences**

| SEQ ID NO | Description | Sequence |
|---|---|---|
| 98 | *Streptococcus pyogenes* Cas9 (spCas9) | MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNT<br>DRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKN<br>RICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHP<br>IFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYL<br>ALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQL<br>FEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKN<br>GLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDD<br>LDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITK<br>APLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFD<br>QSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVK<br>LNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPF<br>LKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEE<br>TITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPK<br>HSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAI<br>VDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRF<br>NASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFE<br>DREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSR<br>KLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTF<br>KEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKV<br>VDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRER<br>MKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNG<br>RDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLT<br>RSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQR<br>KFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQI<br>LDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFY<br>KVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYG<br>DYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEI<br>TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM<br>PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDP<br>KKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGI<br>TIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELE<br>NGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKL<br>KGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADA<br>NLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAF<br>KYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQL<br>GGD |

| | | |
|---|---|---|
| 99 | *Streptococcus pyogenes* Cas9 (SpCas9) D10A | MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNT<br>DRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKN<br>RICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHP<br>IFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYA<br>LAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLF<br>EENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNG<br>LFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDL<br>DNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKA<br>PLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQ<br>SKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKL<br>NREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFL<br>KDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETI<br>TPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKH<br>SLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIV<br>DLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFN<br>ASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFED<br>REMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRK<br>LINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFK<br>EDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVV<br>DELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERM<br>KRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGR<br>DMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTR<br>SDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRK<br>FDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQIL<br>DSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYK<br>VREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGD<br>YKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEIT<br>LANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMP<br>QVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPK<br>KYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITI<br>MERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELEN<br>GRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLK<br>GSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADAN<br>LDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFK<br>YFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLG<br>GD |
| 100 | *Streptococcus pyogenes* Cas9 (SpCas9) H840A | MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNT<br>DRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKN<br>RICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHP<br>IFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYL<br>ALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQL<br>FEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKN<br>GLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDD<br>LDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITK<br>APLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFD<br>QSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVK |

| | | |
|---|---|---|
| | | LNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPF<br>LKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEE<br>TITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPK<br>HSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAI<br>VDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRF<br>NASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFE<br>DREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSR<br>KLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTF<br>KEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKV<br>VDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRER<br>MKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNG<br>RDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLT<br>RSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQR<br>KFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQI<br>LDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFY<br>KVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYG<br>DYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEI<br>TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM<br>PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDP<br>KKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGI<br>TIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELE<br>NGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKL<br>KGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADA<br>NLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAF<br>KYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQL<br>GGD |
| 101 | *Streptococcus pyogenes* Cas9 (SpCas9) D10A and H840A | MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNT<br>DRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKN<br>RICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHP<br>IFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYL<br>ALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQL<br>FEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKN<br>GLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDD<br>LDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITK<br>APLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFD<br>QSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVK<br>LNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPF<br>LKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEE<br>TITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPK<br>HSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAI<br>VDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRF<br>NASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFE<br>DREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSR<br>KLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTF<br>KEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKV<br>VDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRER |

| | | |
|---|---|---|
| | | MKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNG<br>RDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLT<br>RSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQR<br>KFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQI<br>LDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFY<br>KVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYG<br>DYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEI<br>TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM<br>PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDP<br>KKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGI<br>TIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELE<br>NGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKL<br>KGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADA<br>NLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAF<br>KYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQL<br>GGD |
| 102 | eSpCas9(1.1); nCas9 with K848A/K1003A/R1060A | MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNT<br>DRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKN<br>RICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHP<br>IFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYL<br>ALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQL<br>FEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKN<br>GLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDD<br>LDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITK<br>APLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFD<br>QSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVK<br>LNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPF<br>LKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEE<br>TITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPK<br>HSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAI<br>VDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRF<br>NASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFE<br>DREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSR<br>KLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTF<br>KEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKV<br>VDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRER<br>MKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNG<br>RDMYVDQELDINRLSDYDVDHIVPQSFLADDSIDNKVLT<br>RSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQR<br>KFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQI<br>LDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFY<br>KVREINNYHHAHDAYLNAVVGTALIKKYPALESEFVYG<br>DYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEI<br>TLANGEIRKAPLIETNGETGEIVWDKGRDFATVRKVLSM<br>PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDP<br>KKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGI |

| | | |
|---|---|---|
| | | TIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELE<br>NGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKL<br>KGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADA<br>NLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAF<br>KYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQL<br>GGD |
| 103 | nSpCas9-HF1 | DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDR<br>HSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRIC<br>YLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFG<br>NIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALA<br>HMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE<br>NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLF<br>GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDN<br>LLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPL<br>SASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSK<br>NGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNR<br>EDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKD<br>NREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITP<br>WNFEEVVDKGASAQSFIERMTAFDKNLPNEKVLPKHSL<br>LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDL<br>LFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNAS<br>LGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDRE<br>MIEERLKTYAHLFDDKVMKQLKRRRYTGWGALSRKLIN<br>GIRDKQSGKTILDFLKSDGFANRNFMALIHDDSLTFKEDI<br>QKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDEL<br>VKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIE<br>EGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMY<br>VDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDK<br>NRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDN<br>LTKAERGGLSELDKAGFIKRQLVETRAITKHVAQILDSR<br>MNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVRE<br>INNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKV<br>YDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLAN<br>GEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVN<br>IVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYG<br>GFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMER<br>SSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRK<br>RMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSP<br>EDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK<br>VLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDT<br>TIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD |
| 104 | *Staphylococcus aureus* Cas9 | MKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEAN<br>VENNEGRRSKRGARRLKRRRRHRIQRVKKLLFDYNLLT<br>DHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRG<br>VHNVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLE |

| | | |
|---|---|---|
| | | RLKKDGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLD<br>QSFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEML<br>MGHCTYFPEELRSVKYAYNADLYNALNDLNNLVITRDE<br>NEKLEYYEKFQIIENVFKQKKKPTLKQIAKEILVNEEDIK<br>GYRVTSTGKPEFTNLKVYHDIKDITARKEIIENAELLDQI<br>AKILTIYQSSEDIQEELTNLNSELTQEEIEQISNLKGYTGT<br>HNLSLKAINLILDELWHTNDNQIAIFNRLKLVPKKVDLS<br>QQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPND<br>IIIELAREKNSKDAQKMINEMQKRNRQTNERIEEIIRTTGK<br>ENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNNPFNY<br>EVDHIIPRSVSFDNSFNNKVLVKQEENSKKGNRTPFQYLS<br>SSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEERDIN<br>RFSVQKDFINRNLVDTRYATRGLMNLLRSYFRVNNLDV<br>KVKSINGGFTSFLRRKWKFKKERNKGYKHHAEDALIIAN<br>ADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIETEQ<br>EYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRELINDTL<br>YSTRKDDKGNTLIVNNLNGLYDKDNDKLKKLINKSPEK<br>LLMYHHDPQTYQKLKLIMEQYGDEKNPLYKYYEETGN<br>YLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSR<br>NKVVKLSLKPYRFDVYLDNGVYKFVTVKNLDVIKKENY<br>YEVNSKCYEEAKKLKKISNQAEFIASFYNNDLIKINGELY<br>RVIGVNNDLLNRIEVNMIDITYREYLENMNDKRPPRIIKT<br>IASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG |
| 105 | *Francisella tularensis* Cpf1 (Cas12a) | MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDD<br>EKRAKDYKKAKQIIDKYHQFFIEEILSSVCISEDLLQNYS<br>DVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKF<br>KNLFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSD<br>ITDIDEALEIIKSFKGWTTYFKGFHENRKNVYSSNDIPTSII<br>YRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKKDLA<br>EELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGIT<br>KFNTIIGGKFVNGENTKRKGINEYINLYSQQINDKTLKKY<br>KMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSFYEQI<br>AAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSL<br>TDLSQQVFDDYSVIGTAVLEYITQQIAPKNLDNPSKKEQ<br>ELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQCRFEEI<br>LANFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQA<br>SAEDDVKAIKDLLDQTNNLLHKLKIFHISQSEDKANILDK<br>DEHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFK<br>LNFENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNK<br>KNNKIFDDKAIKENKGEGYKKIVYKLLPGANKMLPKVF<br>FSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGYEKFEFNI<br>EDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFY<br>REVENQGYKLTFENISESYIDSVVNQGKLYLFQIYNKDFS<br>AYSKGRPNLHTLYWKALFDERNLQDVVYKLNGEAELF<br>YRKQSIPKKITHPAKEAIANKNKDNPKKESVFEYDLIKD<br>KRFTEDKFFFHCPITINFKSSGANKFNDEINLLLKEKAND |

| | | |
|---|---|---|
| | | VHILSIDRGERHLAYYTLVDGKGNIIKQDTFNIIGNDRMK TNYHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQV VHEIAKLVIEYNAIVVFEDLNFGFKRGRFKVEKQVYQKL EKMLIEKLNYLVFKDNEFDKTGGVLRAYQLTAPFETFK KMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKYESVSK SQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKGKW TIASFGSRLINFRNSDKNHNWDTREVYPTKELEKLLKDY SIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRNSK TGTELDYLISPVADVNGNFFDSRQAPKNMPQDADANGA YHIGLKGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQNR NN |
| 106 | *Acidaminococcus* sp. BV3L6 type V CRISPR-associated protein Cpf1 (Cas12a) | TQFEGFTNLYQVSKTLRFELIPQGKTLKHIQEQGFIEEDK ARNDHYKELKPIIDRIYKTYADQCLQLVQLDWENLSAAI DSYRKEKTEETRNALIEEQATYRNAIHDYFIGRTDNLTD AINKRHAEIYKGLFKAELFNGKVLKQLGTVTTTEHENAL LRSFDKFTTYFSGFYENRKNVFSAEDISTAIPHRIVQDNFP KFKENCHIFTRLITAVPSLREHFENVKKAIGIFVSTSIEEVF SFPFYNQLLTQTQIDLYNQLLGGISREAGTEKIKGLNEVL NLAIQKNDETAHIIASLPHRFIPLFKQILSDRNTLSFILEEF KSDEEVIQSFCKYKTLLRNENVLETAEALFNELNSIDLTH IFISHKKLETISSALCDHWDTLRNALYERRISELTGKITKS AKEKVQRSLKHEDINLQEIISAAGKELSEAFKQKTSEILS HAHAALDQPLPTTLKKQEEKEILKSQLDSLLGLYHLLDW FAVDESNEVDPEFSARLTGIKLEMEPSLSFYNKARNYAT KKPYSVEKFKLNFQMPTLASGWDVNKEKNNGAILFVKN GLYYLGIMPKQKGRYKALSFEPTEKTSEGFDKMYYDYF PDAAKMIPKCSTQLKAVTAHFQTHTTPILLSNNFIEPLEIT KEIYDLNNPEKEPKKFQTAYAKKTGDQKGYREALCKWI DFTRDFLSKYTKTTSIDLSSLRPSSQYKDLGEYYAELNPL LYHISFQRIAEKEIMDAVETGKLYLFQIYNKDFAKGHHG KPNLHTLYWTGLFSPENLAKTSIKLNGQAELFYRPKSRM KRMAHRLGEKMLNKKLKDQKTPIPDTLYQELYDYVNH RLSHDLSDEARALLPNVITKEVSHEIIKDRRFTSDKFFFH VPITLNYQAANSPSKFNQRVNAYLKEHPETPIIGIDRGER NLIYITVIDSTGKILEQRSLNTIQQFDYQKKLDNREKERV AARQAWSVVGTIKDLKQGYLSQVIHEIVDLMIHYQAVV VLENLNFGFKSKRTGIAEKAVYQQFEKMLIDKLNCLVLK DYPAEKVGGVLNPYQLTDQFTSFAKMGTQSGFLFYVPA PYTSKIDPLTGFVDPFVWKTIKNHESRKHFLEGFDFLHY DVKTGDFILHFKMNRNLSFQRGLPGFMPAWDIVFEKNE TQFDAKGTPFIAGKRIVPVIENHRFTGRYRDLYPANELIA LLEEKGIVFRDGSNILPKLLENDDSHAIDTMVALIRSVLQ MRNSNAATGEDYINSPVRDLNGVCFDSRFQNPEWPMDA DANGAYHIALKGQLLLNHLKESKDLKLQNGISNQDWLA YIQELRN |

| | | |
|---|---|---|
| 107 | Cpf1 (AsCpf1 R1225A) | TQFEGFTNLYQVSKTLRFELIPQGKTLKHIQEQGFIEEDK ARNDHYKELKPIIDRIYKTYADQCLQLVQLDWENLSAAI DSYRKEKTEETRNALIEEQATYRNAIHDYFIGRTDNLTD AINKRHAEIYKGLFKAELFNGKVLKQLGTVTTTEHENAL LRSFDKFTTYFSGFYENRKNVFSAEDISTAIPHRIVQDNFP KFKENCHIFTRLITAVPSLREHFENVKKAIGIFVSTSIEEVF SFPFYNQLLTQTQIDLYNQLLGGISREAGTEKIKGLNEVL NLAIQKNDETAHIIASLPHRFIPLFKQILSDRNTLSFILEEF KSDEEVIQSFCKYKTLLRNENVLETAEALFNELNSIDLTH IFISHKKLETISSALCDHWDTLRNALYERRISELTGKITKS AKEKVQRSLKHEDINLQEIISAAGKELSEAFKQKTSEILS HAHAALDQPLPTTLKKQEEKEILKSQLDSLLGLYHLLDW FAVDESNEVDPEFSARLTGIKLEMEPSLSFYNKARNYAT KKPYSVEKFKLNFQMPTLASGWDVNKEKNNGAILFVKN GLYYLGIMPKQKGRYKALSFEPTEKTSEGFDKMYYDYF PDAAKMIPKCSTQLKAVTAHFQTHTTPILLSNNFIEPLEIT KEIYDLNNPEKEPKKFQTAYAKKTGDQKGYREALCKWI DFTRDFLSKYTKTTSIDLSSLRPSSQYKDLGEYYAELNPL LYHISFQRIAEKEIMDAVETGKLYLFQIYNKDFAKGHHG KPNLHTLYWTGLFSPENLAKTSIKLNGQAELFYRPKSRM KRMAHRLGEKMLNKKLKDQKTPIPDTLYQELYDYVNH RLSHDLSDEARALLPNVITKEVSHEIIKDRRFTSDKFFFH VPITLNYQAANSPSKFNQRVNAYLKEHPETPIIGIDRGER NLIYITVIDSTGKILEQRSLNTIQQFDYQKKLDNREKERV AARQAWSVVGTIKDLKQGYLSQVIHEIVDLMIHYQAVV VLENLNFGFKSKRTGIAEKAVYQQFEKMLIDKLNCLVLK DYPAEKVGGVLNPYQLTDQFTSFAKMGTQSGFLFYVPA PYTSKIDPLTGFVDPFVWKTIKNHESRKHFLEGFDFLHY DVKTGDFILHFKMNRNLSFQRGLPGFMPAWDIVFEKNE TQFDAKGTPFIAGKRIVPVIENHRFTGRYRDLYPANELIA LLEEKGIVFRDGSNILPKLLENDDSHAIDTMVALIRSVLQ MANSNAATGEDYINSPVRDLNGVCFDSRFQNPEWPMDA DANGAYHIALKGQLLLNHLKESKDLKLQNGISNQDWLA YIQELRN |
| 108 | *Listeria seeligeri* (WP_01298 5477.1) | MWISIKTLIHHLGVLFFCDYMYNRREKKIIEVKTMRITKV EVDRKKVLISRDKNGGKLVYENEMQDNTEQIMHHKKSS FYKSVVNKTICRPEQKQMKKLVHGLLQENSQEKIKVSD VTKLNISNFLNHRFKKSLYYFPENSPDKSEEYRIEINLSQL LEDSLKKQQGTFICWESFSKDMELYINWAENYISSKTKLI KKSIRNNRIQSTESRSGQLMDRYMKDILNKNKPFDIQSVS EKYQLEKLTSALKATFKEAKKNDKEINYKLKSTLQNHE RQIIEELKENSELNQFNIEIRKHLETYFPIKKTNRKVGDIR NLEIGEIQKIVNHRLKNKIVQRILQEGKLASYEIESTVNSN SLQKIKIEEAFALKFINACLFASNNLRNMVYPVCKKDIL MIGEFKNSFKEIKHKKFIRQWSQFFSQEITVDDIELASWG LRGAIAPIRNEIIHLKKHSWKKFFNNPTFKVKKSKIINGKT |

| | | |
|---|---|---|
| | | KDVTSEFLYKETLFKDYFYSELDSVPELIINKMESSKILD<br>YYSSDQLNQVFTIPNFELSLLTSAVPFAPSFKRVYLKGFD<br>YQNQDEAQPDYNLKLNIYNEKAFNSEAFQAQYSLFKMV<br>YYQVFLPQFTTNNDLFKSSVDFILTLNKERKGYAKAFQD<br>IRKMNKDEKPSEYMSYIQSQLMLYQKKQEEKEKINHFE<br>KFINQVFIKGFNSFIEKNRLTYICHPTKNTVPENDNIEIPFH<br>TDMDDSNIAFWLMCKLLDAKQLSELRNEMIKFSCSLQST<br>EEISTFTKAREVIGLALLNGEKGCNDWKELFDDKEAWK<br>KNMSLYVSEELLQSLPYTQEDGQTPVINRSIDLVKKYGT<br>ETILEKLFSSSDDYKVSAKDIAKLHEYDVTEKIAQQESLH<br>KQWIEKPGLARDSAWTKKYQNVINDISNYQWAKTKVE<br>LTQVRHLHQLTIDLLSRLAGYMSIADRDFQFSSNYILERE<br>NSEYRVTSWILLSENKNKNKYNDYELYNLKNASIKVSSK<br>NDPQLKVDLKQLRLTLEYLELFDNRLKEKRNNISHFNYL<br>NGQLGNSILELFDDARDVLSYDRKLKNAVSKSLKEILSS<br>HGMEVTFKPLYQTNHHLKIDKLQPKKIHHLGEKSTVSSN<br>QVSNEYCQLVRTLLTMK |
| 109 | *Leptotrichia buccalis* (WP_015770004.1) | MKVTKVGGISHKKYTSEGRLVKSESEENRTDERLSALLN<br>MRLDMYIKNPSSTETKENQKRIGKLKKFFSNKMVYLKD<br>NTLSLKNGKKENIDREYSETDILESDVRDKKNFAVLKKI<br>YLNENVNSEELEVFRNDIKKKLNKINSLKYSFEKNKANY<br>QKINENNIEKVEGKSKRNIIYDYYRESAKRDAYVSNVKE<br>AFDKLYKEEDIAKLVLEIENLTKLEKYKIREFYHEIIGRK<br>NDKENFAKIIYEEIQNVNNMKELIEKVPDMSELKKSQVF<br>YKYYLDKEELNDKNIKYAFCHFVEIEMSQLLKNYVYKR<br>LSNISNDKIKRIFEYQNLKKLIENKLLNKLDTYVRNCGKY<br>NYYLQDGEIATSDFIARNRQNEAFLRNIIGVSSVAYFSLR<br>NILETENENDITGRMRGKTVKNNKGEEKYVSGEVDKIY<br>NENKKNEVKENLKMFYSYDFNMDNKNEIEDFFANIDEA<br>ISSIRHGIVHFNLELEGKDIFAFKNIAPSEISKKMFQNEINE<br>KKLKLKIFRQLNSANVFRYLEKYKILNYLKRTRFEFVNK<br>NIPFVPSFTKLYSRIDDLKNSLGIYWKTPKTNDDNKTKEII<br>DAQIYLLKNIYYGEFLNYFMSNNGNFFEISKEIIELNKND<br>KRNLKTGFYKLQKFEDIQEKIPKEYLANIQSLYMINAGN<br>QDEEEKDTYIDFIQKIFLKGFMTYLANNGRLSLIYIGSDE<br>ETNTSLAEKKQEFDKFLKKYEQNNNIKIPYEINEFLREIK<br>LGNILKYTERLNMFYLILKLLNHKELTNLKGSLEKYQSA<br>NKEEAFSDQLELINLLNLDNNRVTEDFELEADEIGKFLDF<br>NGNKVKDNKELKKFDTNKIYFDGENIIKHRAFYNIKKYG<br>MLNLLEKIADKAGYKISIEELKKYSNKKNEIEKNHKMQE<br>NLHRKYARPRKDEKFTDEDYESYKQAIENIEEYTHLKNK<br>VEFNELNLLQGLLLRILHRLVGYTSIWERDLRFRLKGEFP<br>ENQYIEEIFNFENKKNVKYKGGQIVEKYIKFYKELHQND<br>EVKINKYSSANIKVLKQEKKDLYIRNYIAHFNYIPHAEIS<br>LLEVLENLRKLLSYDRKLKNAVMKSVVDILKEYGFVAT |

| | | |
|---|---|---|
| | | FKIGADKKIGIQTLESEKIVHLKNLKKKKLMTDRNSEELC KLVKIMFEYKMEEKKSEN |
| 110 | *Leptotrichia shahii* (WP_018451595.1) | MGNLFGHKRWYEVRDKKDFKIKRKVKVKRNYDGNKYI LNINENNNKEKIDNNKFIRKYINYKKNDNILKEFTRKFHA GNILFKLKGKEGIIRIENNDDFLETEEVVLYIEAYGKSEKL KALGITKKKIIDEAIRQGITKDDKKIEIKRQENEEEIEIDIR DEYTNKTLNDCSIILRIIENDELETKKSIYEIFKNINMSLYK IIEKIIENETEKVFENRYYEEHLREKLLKDDKIDVILTNFM EIREKIKSNLEILGFVKFYLNVGGDKKKSKNKKMLVEKI LNINVDLTVEDIADFVIKELEFWNITKRIEKVKKVNNEFL EKRRNRTYIKSYVLLDKHEKFKIERENKKDKIVKFFVE NIKNNSIKEKIEKILAEFKIDELIKKLEKELKKGNCDTEIF GIFKKHYKVNFDSKKFSKKSDEEKELYKIIYRYLKGRIEK ILVNEQKVRLKKMEKIEIEKILNESILSEKILKRVKQYTLE HIMYLGKLRHNDIDMTTVNTDDFSRLHAKEELDLELITF FASTNMELNKIFSRENINNDENIDFFGGDREKNYVLDKKI LNSKIKIIRDLDFIDNKNNITNNFIRKFTKIGTNERNRILHA ISKERDLQGTQDDYNKVINIIQNLKISDEEVSKALNLDVV FKDKKNIITKINDIKISEENNNDIKYLPSFSKVLPEILNLYR NNPKNEPFDTIETEKIVLNALIYVNKELYKKLILEDDLEE NESKNIFLQELKKTLGNIDEIDENIIENYYKNAQISASKGN NKAIKKYQKKVIECYIGYLRKNYEELFDFSDFKMNIQEIK KQIKDINDNKTYERITVKTSDKTIVINDDFEYIISIFALLNS NAVINKIRNRFFATSVWLNTSEYQNIIDILDEIMQLNTLR NECITENWNLNLEEFIQKMKEIEKDFDDFKIQTKKEIFNN YYEDIKNNILTEFKDDINGCDVLEKKLEKIVIFDDETKFEI DKKSNILQDEQRKLSNINKKDLKKKVDQYIKDKDQEIKS KILCRIIFNSDFLKKYKKEIDNLIEDMESENENKFQEIYYP KERKNELYIYKKNLFLNIGNPNFDKIYGLISNDIKMADA KFLFNIDGKNIRKNKISEIDAILKNLNDKLNGYSKEYKEK YIKKLKENDDFFAKNIQNKNYKSFEKDYNRVSEYKKIRD LVEFNYLNKIESYLIDINWKLAIQMARFERDMHYIVNGL RELGIIKLSGYNTGISRAYPKRNGSDGFYTTTAYYKFFDE ESYKKFEKICYGFGIDLSENSEINKPENESIRNYISHFYIVR NPFADYSIAEQIDRVSNLLSYSTRYNNSTYASVFEVFKKD VNLDYDELKKKFKLIGNNDILERLMKPKKVSVLELESYN SDYIKNLIIELLTKIENTNDTL |
| 111 | *Rhodobacter capsulatus* (R121) (ETD76934.1) (U717_11515) | MQIGKVQGRTISEFGDPAGGLKRKISTDGKNRKELPAHL SSDPKALIGQWISGIDKIYRKPDSRKSDGKAIHSPTPSKM QFDARDDLGEAFWKLVSEAGLAQDSDYDQFKRRLHPY GDKFQPADSGAKLKFEADPPEPQAFHGRWYGAMSKRG NDAKELAAALYEHLHVDEKRIDGQPKRNPKTDKFAPGL VVARALGIESSVLPRGMARLARNWGEEEIQTYFVVDVA ASVKEVAKAAVSAAQAFDPPRQVSGRSLSPKVGFALAE HLERVTGSKRCSFDPAAGPSVLALHDEVKKTYKRLCAR |

| | | |
|---|---|---|
| | | GKNAARAFPADKTELLALMRHTHENRVRNQMVRMGR<br>VSEYRGQQAGDLAQSHYWTSAGQTEIKESEIFVRLWVG<br>AFALAGRSMKAWIDPMGKIVNTEKNDRDLTAAVNIRQV<br>ISNKEMVAEAMARRGIYFGETPELDRLGAEGNEGFVFAL<br>LRYLRGCRNQTFHLGARAGFLKEIRKELEKTRWGKAKE<br>AEHVVLTDKTVAAIRAIIDNDAKALGARLLADLSGAFVA<br>HYASKEHFSTLYSEIVKAVKDAPEVSSGLPRLKLLLKRA<br>DGVRGYVHGLRDTRKHAFATKLPPPPAPRELDDPATKA<br>RYIALLRLYDGPFRAYASGITGTALAGPAARAKEAATAL<br>AQSVNVTKAYSDVMEGRSSRLRPPNDGETLREYLSALT<br>GETATEFRVQIGYESDSENARKQAEFIENYRRDMLAFMF<br>EDYIRAKGFDWILKIEPGATAMTRAPVLPEPIDTRGQYE<br>HWQAALYLVMHFVPASDVSNLLHQLRKWEALQGKYEL<br>VQDGDATDQADARREALDLVKRFRDVLVLFLKTGEARF<br>EGRAAPFDLKPFRALFANPATFDRLFMATPTTARPAEDD<br>PEGDGASEPELRVARTLRGLRQIARYNHMAVLSDLFAK<br>HKVRDEEVARLAEIEDETQEKSQIVAAQELRTDLHDKV<br>MKCHPKTISPEERQSYAAAIKTIEEHRFLVGRVYLGDHL<br>RLHRLMMDVIGRLIDYAGAYERDTGTFLINASKQLGAG<br>ADWAVTIAGAANTDARTQTRKDLAHFNVLDRADGTPD<br>LTALVNRAREMMAYDRKRKNAVPRSILDMLARLGLTL<br>KWQMKDHLLQDATITQAAIKHLDKVRLTVGGPAAVTE<br>ARFSQDYLQMVAAVFNGSVQNPKPRRRDDGDAWHKPP<br>KPATAQSQPDQKPPNKAPSAGSRLPPPQVGEVYEGVVV<br>KVIDTGSLGFLAVEGVAGNIGLHISRLRRIREDAIIVGRRY<br>RFRVEIYVPPKSNTSKLNAADLVRID |
| 112 | WP_034560163.1 WP_034560163.1 hypothetical protein [*Carnobacterium gallinarum*] | MRITKVKIKLDNKLYQVTMQKEEKYGTLKLNEESRKST<br>AEILRLKKASFNKSFHSKTINSQKENKNATIKKNGDYISQ<br>IFEKLVGVDTNKNIRKPKMSLTDLKDLPKKDLALFIKRK<br>FKNDDIVEIKNLDLISLFYNALQKVPGEHFTDESWADFC<br>QEMMPYREYKNKFIERKIILLANSIEQNKGFSINPETFSKR<br>KRVLHQWAIEVQERGDFSILDEKLSKLAEIYNFKKMCKR<br>VQDELNDLEKSMKKGKNPEKEKEAYKKQKNFKIKTIWK<br>DYPYKTHIGLIEKIKENEELNQFNIEIGKYFEHYFPIKKER<br>CTEDEPYYLNSETIATTVNYQLKNALISYLMQIGKYKQF<br>GLENQVLDSKKLQEIGIYEGFQTKFMDACVFATSSLKNII<br>EPMRSGDILGKREFKEAIATSSFVNYHHFFPYFPFELKGM<br>KDRESELIPFGEQTEAKQMQNIWALRGSVQQIRNEIFHSF<br>DKNQKFNLPQLDKSNFEFDASENSTGKSQSYIETDYKFL<br>FEAEKNQLEQFFIERIKSSGALEYYPLKSLEKLFAKKEMK<br>FSLGSQVVAFAPSYKKLVKKGHSYQTATEGTANYLGLS<br>YYNRYELKEESFQAQYYLLKLIYQYVFLPNFSQGNSPAF<br>RETVKAILRINKDEARKKMKKNKKFLRKYAFEQVREME<br>FKETPDQYMSYLQSEMREEKVRKAEKNDKGFEKNITMN<br>FEKLLMQIFVKGFDVFLTTFAGKELLLSSEEKVIKETEISL<br>SKKINEREKTLKASIQVEHQLVATNSAISYWLFCKLLDSR |

| | | |
|---|---|---|
| | | HLNELRNEMIKFKQSRIKFNHTQHAELIQNLLPIVELTILS<br>NDYDEKNDSQNVDVSAYFEDKSLYETAPYVQTDDRTR<br>VSFRPILKLEKYHTKSLIEALLKDNPQFRVAATDIQEWM<br>HKREEIGELVEKRKNLHTEWAEGQQTLGAEKREEYRDY<br>CKKIDRFNWKANKVTLTYLSQLHYLITDLLGRMVGFSA<br>LFERDLVYFSRSFSELGGETYHISDYKNLSGVLRLNAEV<br>KPIKIKNIKVIDNEENPYKGNEPEVKPFLDRLHAYLENVI<br>GIKAVHGKIRNQTAHLSVLQLELSMIESMNNLRDLMAY<br>DRKLKNAVTKSMIKILDKHGMILKLKIDENHKNFEIESLI<br>PKEIIHLKDKAIKTNQVSEEYCQLVLALLTTNPGNQLN |
| 113 | CRZ35554.1 *Herbinix hemicellulosilytica* genome assembly TUM3/55, contig 02_T3/55T_contig26, whole genome shotgun sequence | MKLTRRRISGNSVDQKITAAFYRDMSQGLLYYDSEDND<br>CTDKVIESMDFERSWRGRILKNGEDDKNPFYMFVKGLV<br>GSNDKIVCEPIDVDSDPDNLDILINKNLTGFGRNLKAPDS<br>NDTLENLIRKIQAGIPEEEVLPELKKIKEMIQKDIVNRKEQ<br>LLKSIKNNRIPFSLEGSKLVPSTKKMKWLFKLIDVPNKTF<br>NEKMLEKYWEIYDYDKLKANITNRLDKTDKKARSISRA<br>VSEELREYHKNLRTNYNRFVSGDRPAAGLDNGGSAKYN<br>PDKEEFLLFLKEVEQYFKKYFPVKSKHSNKSKDKSLVDK<br>YKNYCSYKVVKKEVNRSIINQLVAGLIQQGKLLYYFYY<br>NDTWQEDFLNSYGLSYIQVEEAFKKSVMTSLSWGINRLT<br>SFFIDDSNTVKFDDITTKKAKEAIESNYFNKLRTCSRMQD<br>HFKEKLAFFYPVYVKDKKDRPDDDIENLIVLVKNAIESV<br>SYLRNRTFHFKESSLLELLKELDDKNSGQNKIDYSVAAE<br>FIKRDIENLYDVFREQIRSLGIAEYYKADMISDCFKTCGL<br>EFALYSPKNSLMPAFKNVYKRGANLNKAYIRDKGPKET<br>GDQGQNSYKALEEYRELTWYIEVKNNDQSYNAYKNLL<br>QLIYYHAFLPEVRENEALITDFINRTKEWNRKETEERLNT<br>KNNKKHKNFDENDDITVNTYRYESIPDYQGESLDDYLK<br>VLQRKQMARAKEVNEKEEGNNNYIQFIRDVVVWAFGA<br>YLENKLKNYKNELQPPLSKENIGLNDTLKELFPEEKVKS<br>PFNIKCRFSISTFIDNKGKSTDNTSAEAVKTDGKEDEKDK<br>KNIKRKDLLCFYLFLRLLDENEICKLQHQFIKYRCSLKER<br>RFPGNRTKLEKETELLAELEELMELVRFTMPSIPEISAKA<br>ESGYDTMIKKYFKDFIEKKVFKNPKTSNLYYHSDSKTPV<br>TRKYMALLMRSAPLHLYKDIFKGYYLITKKECLEYIKLS<br>NIIKDYQNSLNELHEQLERIKLKSEKQNGKDSLYLDKKD<br>FYKVKEYVENLEQVARYKHLQHKINFESLYRIFRIHVDI<br>AARMVGYTQDWERDMHFLFKALVYNGVLEERRFEAIF<br>NNNDDNNDGRIVKKIQNNLNNKNRELVSMLCWNKKLN<br>KNEFGAIIWKRNPIAHLNHFTQTEQNSKSSLESLINSLRIL<br>LAYDRKRQNAVTKTINDLLLNDYHIRIKWEGRVDEGQI<br>YFNIKEKEDIENEPIIHLKHLHKKDCYIYKNSYMFDKQKE<br>WICNGIKEEVYDKSILKCIGNLFKFDYEDKNKSSANPKH<br>T |

**[0259]** In some embodiments, the CRISPR-Cas nuclease is a Type II CRISPR-Cas nuclease. In some embodiments, the CRISPR-Cas nuclease is a Cas9 polypeptide. The Cas9 protein is guided to a target site (e.g., stabilized at a target site) within a target nucleic acid sequence (e.g., a chromosomal sequence or an extrachromosomal sequence, e.g., an episomal sequence, a minicircle sequence, a mitochondrial sequence, a chloroplast sequence, etc.) by virtue of its association with the protein-binding segment of the Cas9 gRNA. In some embodiments, a Cas9 polypeptide comprises an amino acid sequence having at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, at least 98%, at least 99%, or more than 99%, amino acid sequence identity to the *Streptococcus pyogenes* Cas9 depicted in SEQ ID NO: 98.

**[0260]** In some embodiments, the Cas9 polypeptide is a *Staphylococcus aureus* Cas9 (saCas9) polypeptide. In some embodiments, the saCas9 polypeptide comprises an amino acid sequence having at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, or 100%, amino acid sequence identity to the saCas9 amino acid sequence depicted in SEQ ID NO: 104.

**[0261]** In some embodiments, a suitable Cas9 polypeptide is a high-fidelity (HF) Cas9 polypeptide. Kleinstiver et al. (2016) Nature 529:490. For example, amino acids N497, R661, Q695, and Q926 of the amino acid sequence depicted in SEQ ID NO: 98 are substituted, e.g., with alanine. For example, an HF Cas9 polypeptide can comprise an amino acid sequence having at least 90%, at least 95%, at least 98%, at least 99%, or 100%, amino acid sequence identity to the amino acid sequence depicted in SEQ ID NO: 98, where amino acids N497, R661, Q695, and Q926 are substituted, e.g., with alanine. In some embodiments, a suitable Cas9 polypeptide exhibits altered PAM specificity. See, e.g., Kleinstiver et al. (2015) Nature 523:481.

**[0262]** In some embodiments, a suitable CRISPR-Cas nuclease is a type V CRISPR-Cas nuclease. In some embodiments, a type V CRISPR-Cas nuclease is a Cpf1 protein. In some embodiments, a Cpf1 protein comprises an amino acid sequence having at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 90%, or 100%, amino acid sequence identity to the Cpf1 amino acid sequence depicted in SEQ ID NO: 105, 106, or 107.

### i. Class 2 CRISPR-Cas Nucleases

**[0263]** In class 2 CRISPR systems, the functions of the effector complex (e.g., the cleavage of target DNA) are carried out by a single endonuclease (e.g., see Zetsche et al., Cell. 2015 Oct 22;163(3):759-71; Makarova et al., Nat Rev Microbiol. 2015 Nov;13(11):722-36; Shmakov et al., Mol Cell. 2015 Nov 5;60(3):385-97); and Shmakov et al. (2017) Nature Reviews Microbiology 15:169. As such, the term "class 2 CRISPR-Cas protein" is used herein to encompass the CRISPR-Cas nuclease (e.g., the target nucleic acid cleaving protein) from class 2 CRISPR systems. Thus, the term "class 2 CRISPR-Cas nuclease" as used herein encompasses type II CRISPR-Cas nucleases (e.g., Cas9); type V-A CRISPR-Cas nucleases (e.g., Cpf1 (also referred to a "Cas12a")); type V-B CRISPR-Cas nucleases (e.g., C2c1 (also referred to as "Cas12b")); type V-C CRISPR-Cas nucleases (e.g., C2c3 (also referred to as "Cas12c")); type V-U1 CRISPR-Cas nucleases (e.g., C2c4); type V-U2 CRISPR-Cas nucleases (e.g., C2c8); type V-U5 CRISPR-Cas nucleases (e.g., C2c5); type V-U4 CRISPR-Cas proteins (e.g., C2c9); type V-U3 CRISPR-Cas nucleases (e.g., C2c10); type VI-A CRISPR-Cas nucleases (e.g., C2c2 (also known as "Cas13a")); type VI-B CRISPR-Cas nucleases (e.g., Cas13b (also known as C2c4)); and type VI-C CRISPR-Cas nucleases (e.g., Cas13c (also known as C2c7)). To date, class 2 CRISPR-Cas nucleases encompass type II, type V, and type VI CRISPR-Cas nucleases, but the term is also meant to encompass any class 2 CRISPR-Cas nuclease suitable for binding to a corresponding gRNA and forming a ribonucleoprotein (RNP) complex.

### ii. Fusion Polypeptides

**[0264]** A polynucleotide-guided nuclease can be covalently linked to one or more heterologous polypeptides (also referred to as a "fusion partner") to form a fusion polypeptide. In some embodiments, the polynucleotide-guided nuclease is a fusion polypeptide comprising: i) a polynucleotide-guided nuclease ; and ii) one or more heterologous partners (one or more heterologous polypeptides). In some embodiments, the polynucleotide-guided nuclease is a CRISPR-Cas nuclease. In some embodiments, the polynucleotide-guided nuclease is a CRISPR-Cas9 nickase, that is optionally covalently liked to (1) a serine integrase or a serine recombinase (i.e., a CRISPR-directed integrase or a CRISPR-directed recombinase) with or without a (2) reverse transcriptase. In some embodiments, the polynucleotide-guided nuclease is a CRISPR-associated transposase (CAST).

**[0265]** In some embodiments, a fusion polynucleotide-guided nuclease comprises one or more localization signal peptides. Suitable localization signals ("subcellular localization signals") include, e.g., a nuclear localization signal (NLS) for targeting to the nucleus; a sequence to keep the fusion protein out of the nucleus, e.g., a nuclear export sequence (NES); a sequence to keep the fusion protein retained in the cytoplasm; a mitochondrial localization signal for targeting to the mitochondria; a chloroplast localization signal for targeting to a chloroplast; an endoplasmic reticulum (ER) retention signal; and ER export signal; and the like. In some embodiments, a fusion CRISPR-Cas nuclease does not include a NLS so that the protein is not targeted to the nucleus (which can be advantageous, e.g., when the target nucleic acid is an RNA that is present in the cytosol). In some embodiments, a fusion CRISPR-Cas nuclease comprises both an NES and one or more NLSs.

**[0266]** In some embodiments, the polynucleotide-guided nuclease, e.g., CRISPR-Cas nuclease, is covalently linked to an NES. A suitable NES comprises hydrophobic amino acid residues, e.g., LXXXLXXLXL (SEQ ID NO: 114), where L is a hydrophobic amino acid residue (e.g., Leu) and X is any other amino acid. Suitable NESs are known in the art; see, e.g., Xu et al. (2012) Mol. Biol. Cell 23:3677. Non-limiting examples of suitable NESs are shown below in Table 4. In some embodiments, the polynucleotide-guided nuclease, e.g., CRISPR-Cas nuclease, is covalently linked to 4 or more, 5 or more, 6 or more, 7 or more, 8 or more, 9 or more, or 10 or more NESs. In some embodiments, the polynucleotide-guided nuclease is linked to the NESs on its N-terminus, C-terminus, or both.

**Table 4: Exemplary Nuclear Export Signal (NES) Peptide Sequences**

| SEQ ID NO | Description | Sequence |
|---|---|---|
| 114 | Exemplary NES; L is a hydrophobic amino acid residue (e.g., Leu) and X is any other amino acid | LXXXLXXLXL |
| 115 | Exemplary NES | LPPLERLTL |

| | | |
|---|---|---|
| 116 | Exemplary NES | LALKLAGLDL |
| 117 | Exemplary NES | LSQALASSFSV |
| 118 | Exemplary NES | NELALKLAGLDI |
| 119 | Exemplary NES | LPPLERLTL |

**[0267]** Polynucleotide-guided nucleases of the present disclosure, e.g., CRISPR-Cas nucleases, are designed to contact a target cell nucleus. In some embodiments, a fusion polynucleotide-guided nuclease is covalently linked to one or more NLSs (e.g., 4 or more, 5 or more, 6 or more, 7 or more, 8 or more, 9 or more, or 10 or more NLSs). In some embodiments, one or more NLSs (4 or more, 5 or more, 6 or more, 7 or more, 8 or more, 9 or more, or 10 or more NLSs) are positioned at or near (e.g., within 50 amino acids of) the N-terminus and/or the C-terminus. In some embodiments, one or more NLSs (4 or more, 5 or more, 6 or more, 7 or more, 8 or more, 9 or more, or 10 or more NLSs) are positioned at or near (e.g., within 50 amino acids of) the N-terminus. In some embodiments, one or more NLSs (4 or more, 5 or more, 6 or more, 7 or more, 8 or more, 9 or more, or 10 or more NLSs) are positioned at or near (e.g., within 50 amino acids of) the C-terminus. In some embodiments, one or more NLSs (3 or more, 4 or more, 5 or more, or 6 or more NLSs) are positioned at or near (e.g., within 50 amino acids of) both the N-terminus and the C-terminus (for a total of 6 or more, 7 or more, 8 or more, 9 or more, 10 or more, 11 or more, or 12 or more NLSs). In some embodiments, the polynucleotide-guided nuclease (e.g., CRISPR-Cas nuclease) is fused to 7 or more NLSs on its C-terminus.

**[0268]** In some embodiments, a CRISPR-Cas fusion polypeptide comprises: a) a CRISPR-Cas nuclease; and b) from 1 to 10 NLSs (e.g., 1-9, 1-8, 1-7, 1-6, 1-5, 2-10, 2-9, 2-8, 2-7, 2-6, or 2-5 NLSs). In some embodiments, a CRISPR-Cas fusion polypeptide comprises: a) a CRISPR-Cas nuclease; and b) 7 NLSs. In some embodiments, the NLS is the SV40 NLS (SEQ ID NO: 120; see Table 7 below).

**[0269]** Non-limiting examples of NLSs include an NLS sequence as shown in Table 5 below. In general, NLS (or multiple NLSs) are of sufficient strength to drive accumulation of the CRISPR-

Cas nuclease in a detectable amount in the nucleus of a eukaryotic cell. Detection of accumulation in the nucleus may be performed by any suitable technique. For example, a detectable marker may be fused to the CRISPR-Cas nuclease such that location within a cell may be visualized. Cell nuclei may also be isolated from cells, the contents of which may then be analyzed by any suitable process for detecting protein, such as immunohistochemistry, Western blot, or enzyme activity assay. Accumulation in the nucleus may also be determined indirectly.

**Table 5: Exemplary Nuclear Localization Signal (NLS) Peptide Sequences**

| SEQ ID NO | Description | Sequence |
|---|---|---|
| 120 | NLS of the SV40 virus large T-antigen | PKKKRKV |
| 121 | NLS from nucleoplasmin, e.g., the nucleoplasmin bipartite NLS | KRPAATKKAGQAKKKK |
| 122 | c-myc NLS | PAAKRVKLD |
| 123 | c-myc NLS | RQRRNELKRSP |
| 124 | hRNPA1 M9 NLS | NQSSNFGPMKGGNFGGRSSGPYGGGGQYFAKPRNQGGY |
| 125 | Sequence from the IBB domain from importin-α | RMRIZFKNKGKDTAELRRRRVEVSVELRKAKKDEQILKRRNV |
| 126 | Sequence from myoma T protein | VSRKRPRP |
| 127 | Sequence from myoma T protein | PPKKARED |

| | | |
|---|---|---|
| 128 | Sequence from human p53 | PQPKKKPL |
| 129 | Sequence frommouse c-abl IV | SALIKKKKKMAP |
| 130 | Sequence from influenza virus NS1 | DRLRR |
| 131 | Sequence from influenza virus NS1 | PKQKKRK |
| 132 | Hepatitis virus delta antigen | RKLKKKIKKL |
| 133 | Sequence from mouse Mx1 protein | REKKKFLKRR |
| 134 | Sequence from human poly(ADP-ribose) polymerase | KRKGDEVDGVDEVAKKKSKK |
| 135 | Sequence from steroid hormone receptors (human) glucocorticoid | RKCLQAGMNLEARKTKK |
| 136 | Exemplary NLS | MDSLLMNRRKFLYQFKNVRWAKGRRETYLC |

**[0270]** In some embodiments, polynucleotide-guided nuclease, e.g., CRISPR-Cas nuclease, is covalently linked to a heterologous polypeptide. In some embodiments, the heterologous polypeptide is a protein modifying enzyme. In some embodiments, the heterologous polypeptide

is a nucleic acid modifying enzyme, e.g., a cytidine deaminase, adenosine deaminase, or prime editor. In some embodiments, the heterologous polypeptide is a reverse transcriptase, a serine recombinase, or both. In some embodiments, the heterologous polypeptide is a reverse transcriptase (e.g., as used in prime editing), a serine integrase, or both. In some embodiments, the heterologous polypeptide is a cytidine deaminase. In some embodiments, the heterologous polypeptide is an adenine deaminase. In some embodiments, the heterologous polypeptide is a transcription factor. In some embodiments, the heterologous polypeptide is a transcription activator. In some embodiments, the heterologous polypeptide is a transcription repressor. In some embodiments, the heterologous polypeptide is or comprises a serine recombinase or a serine integrase. In some embodiments, the heterologous polypeptide is or comprises a transposase domain.

**[0271]** In some embodiments, the heterologous polypeptide is a reverse transcriptase covalently linked to a serine integrase or a serine recombinase. In some embodiments, the polynucleotide-guided nuclease is a CRISPR-Cas9 nickase, and the CRISPR-Cas9 nickase is covalently linked to (1) a serine integrase or a serine recombinase (i.e., a CRISPR-directed integrase or a CRISPR-directed recombinase) with or without a (2) reverse transcriptase. CRISPR-directed integrases and related site-specific targeting elements (PASTE) methods are discussed in Yarnall, Matthew T N et al. "Drag-and-drop genome insertion of large sequences without double-strand DNA cleavage using CRISPR-directed integrases." Nature biotechnology vol. 41,4 (2023): 500-512. doi:10.1038/s41587-022-01527-4.

**[0272]** In some embodiments, the polynucleotide-guided nuclease is a CRISPR-Cas nuclease covalently linked to a transposase, i.e., a CRISPR-associated transposase (CAST). In these embodiments, an insertion sequence can be introduced to a target locus without reliance on a donor template polynucleotide, homologous recombination, or double-stranded breaks. In these embodiments, the CAST is directed to the target locus by the gRNA. CASTs and methods of using them are discussed in Lampe, George D et al. "Targeted DNA integration in human cells without double-strand breaks using CRISPR-associated transposases." Nature biotechnology vol. 42,1 (2024): 87-98. doi:10.1038/s41587-023-01748-1; and Tou, Connor J et al. "Precise cut-and-paste DNA insertion using engineered type V-K CRISPR-associated transposases." Nature biotechnology vol. 41,7 (2023): 968-979. doi:10.1038/s41587-022-01574-x; Trujillo Rodríguez,

Lidimarie et al. "CRISPR-Associated Transposase for Targeted Mutagenesis in Diverse Proteobacteria." ACS synthetic biology vol. 12,7 (2023): 1989-2003. doi:10.1021/acssynbio.3c00065.

**b. Guide Polynucleotides**

**[0273]** A guide polynucleotide is a nucleic acid molecule that can bind to a polynucleotide-guided nuclease, thereby forming a nucleic acid-protein complex, and can target the complex to a specific location within a target nucleic acid (e.g., target locus on genomic DNA). In some embodiments, the polynucleotide-guided nuclease is a CRISPR-Cas nuclease, the guide polynucleotide is a "CRISPR-Cas nuclease guide RNA" or simply a "guide RNA," and the nucleic acid-protein complex is a ribonucleoprotein (RNP) complex.

**[0274]** As discussed above, an EDV of the present disclosure may comprise a guide polynucleotide that directs a polynucleotide-guided nuclease to a specific site for cleaving genomic DNA. In some embodiments, the EDV comprises an CRISPR-Cas nuclease and a guide RNA (gRNA). In some embodiments, instead of a gRNA, the EDV comprises a nucleic acid comprising a nucleotide sequence encoding a gRNA.

**[0275]** In some embodiments, the EDV comprises an CRISPR-Cas nuclease and two or more gRNAs, where the two or more gRNAs provide for multiplexed gene modifications. For example, each of the two or more gRNAs is targeted to a different locus or a different gene.

**[0276]** In some embodiments, the AAV comprises one or more gRNAs instead of or in addition to the EDV-comprised gRNAs.

**[0277]** In some embodiments, the EDV comprises a non-targeting gRNA that stabilizes the RNP and the AAV comprises a gRNA for directing a polynucleotide-guided nuclease to a specific DNA locus, e.g., including but not limited to a locus as described herein.

**[0278]** In some embodiments, the AAV comprises an additional gRNA that directs the nuclease to disrupt a gene (e.g., knockout without a donor template added to the disrupted locus), wherein the gene disruption will promote, enhance, or synergize with the therapeutic (e.g., anti-cancer) activity of the heterologous polypeptide encoded by the donor template polynucleotide. In some embodiments of these embodiments, the additional gRNA is targeted to, for example, A20

(TNFAIP3), PD1, CTLA4, Roquin, REGNASE-1, RASA2, ARID1a, Cul5, SUV39H1, Med12, Med24, ADORA2A, CBLB, CD5, CD7, CISH, DGKa, DGKz, DNMT3A, FAS, FLI1, IKZF3, MAPK14, PTPN2, SMARCD1, SOCS1, TOX, or VHL.

**[0279]** In some embodiments, a guide polynucleotide, e.g., gRNA, comprises a sequence having at least 70%, at least 80%, at least 85%, at least 90%, at least 95%, or 100% identity to a sequence at a target site. In some embodiments, the target site is on the genomic DNA of a target cell. In many embodiments, the guide polynucleotide, e.g., gRNA, targets an exonic region, an intronic region, or both an exonic region and an intronic region of a gene. In some embodiments, the target site is a TCR locus. Insertion of heterologous sequences into a TCR locus is discussed in detail below.

**[0280]** In some embodiments, the target cell is a T cell and the guide polynucleotide, e.g., gRNA, comprises a sequence having at least 70%, at least 80%, at least 85%, at least 90%, at least 95%, or 100% identity to a sequence of any one of T cell receptor α constant (TRAC), T cell receptor beta constant (TRBC), T cell receptor gamma constant (TRGC), T cell receptor delta constant (TRDC), CD4, CD5, CD6, CD7, CD8a, CD8b, CD3e, CD247, CD27, CD28, interleukin-2R (IL-2R) α, IL-2R beta, killer cell lectin-like receptor (KLR) C1 (KLRC1), KLRF1, KLRG1, granzyme (GZM) A (GZMA), GZMB, GZMH, GZMK, Zap-70, lymphocyte-specific protein tyrosine kinase (LCK), linker for activation of T cells (LAT), IL-2 inducible T cell kinase (ITK), transcription factor 7 (TCF7), lymphoid enhancer binding factor 1 (LEF1), or FoxP3.

**[0281]** In some embodiments, the target cell is a B cell and the guide polynucleotide, e.g., gRNA, comprises a sequence having at least 70%, at least 80%, at least 85%, at least 90%, at least 95%, or 100% identity to a sequence of any one of CD19, CD20, CD22, CD138, BCMA, TACI, MS4A1, IGH, IGK, CD79A, or CD79B.

**[0282]** In some embodiments, the target cell is an NK cell and the guide polynucleotide, e.g., gRNA, comprises a sequence having at least 70%, at least 80%, at least 85%, at least 90%, at least 95%, or 100% identity to a sequence of any one of NCAM1, FCGR3A, NCR1, NCR2, KLRC1, NKG2D, NKG7, KIR2DL1, KIR2DL2, KIR2DL3, KIR2DL4, KIR3DL1 or KIR3DL1.

**[0283]** In some embodiments, the target cell is a monocyte or a macrophage and the guide polynucleotide, e.g., gRNA, comprises a sequence having at least 70%, at least 80%, at least 85%,

at least 90%, at least 95%, or 100% identity to a sequence of any one of CD11b, CD11c, CD14, CD33, CD163, CLEC7A, C1QA, C1QB, C1QC, or MSR1.

**[0284]** In some embodiments, the target cell is a dendritic cell and the guide polynucleotide, e.g., gRNA, comprises a sequence having at least 70%, at least 80%, at least 85%, at least 90%, at least 95%, or 100% identity to a sequence of any one of CD1C, DCIR, CLEC10A, NDRG2, or TPM2.

**[0285]** In some embodiments, the target cell is an HSC and the guide polynucleotide, e.g., gRNA, comprises a sequence having at least 70%, at least 80%, at least 85%, at least 90%, at least 95%, or 100% identity to a sequence of any one of PTPRC, CD34, HBB, or RAG2.

### i. Single guide RNA (gRNA)

**[0286]** A gRNA can be said to include two segments: a first segment, referred to herein as a "targeter," a "targeting segment," or a "targeting sequence;" and a second segment, referred to herein as an "activator," a "protein-binding segment" or a "protein-binding sequence." The term "segment" can refer to a segment/section/region/portion of a molecule, e.g., a contiguous stretch of nucleotides in a nucleic acid molecule. A segment can also refer to a region/section/portion of a complex such that a segment may comprise regions of more than one molecule. The "targeting segment" is also referred to herein as a "variable region" of a gRNA. The "protein-binding segment" is also referred to herein as a "constant region" of a gRNA. In some embodiments, the gRNA is a Cas9 gRNA. A gRNA can comprise any corresponding targeter and activator pair.

**[0287]** The first segment, i.e., targeting segment, of the gRNA includes a nucleotide sequence (a guide sequence) that is complementary to (and therefore hybridizes with) a specific sequence (a target site) within a target nucleic acid (e.g., a target DNA, such as a dsDNA and ssDNA; or a target RNA), such as a complementary strand of a double stranded target genomic DNA, etc. The second segment, i.e., protein-binding segment, of the gRNA interacts with (e.g., binds to) a CRISPR-Cas nuclease. The protein-binding segment of a gRNA includes two complementary stretches of nucleotides that hybridize to one another to form a double stranded RNA duplex (dsRNA duplex). Site-specific binding and/or cleavage of a target nucleic acid (e.g., genomic DNA) can occur at locations determined by base-pairing complementarity between the gRNA (i.e.,

the guide sequence of the gRNA) and the target nucleic acid (e.g., a target sequence of a target locus on genomic DNA).

**[0288]** A gRNA and a CRISPR-Cas nuclease form a complex (e.g., bind via non-covalent interactions). The gRNA provides target specificity to the complex by including a targeting segment, which includes a guide sequence (a nucleotide sequence that is complementary to a sequence of a target nucleic acid, e.g., target locus on genomic DNA). The CRISPR-Cas nuclease of the complex provides the site-specific activity (e.g., cleavage activity, or an activity provided by the CRISPR-Cas nuclease when the CRISPR-Cas nuclease is a CRISPR-Cas nuclease fusion polypeptide, i.e., has a fusion partner). CRISPR-Cas nuclease fusion polypeptides and exemplary fusion partners are discussed above in detail. In other words, the CRISPR-Cas nuclease is guided to a target nucleic acid sequence (e.g. a target sequence of a target locus on genomic DNA, e.g., a chromosome; a target sequence in an extrachromosomal nucleic acid; e.g., an episomal nucleic acid, a minicircle, an ssRNA, an ssDNA, etc.; a target sequence in a mitochondrial nucleic acid; a target sequence in a chloroplast nucleic acid; a target sequence in a plasmid; a target sequence in a viral nucleic acid; etc.) by virtue of its association with the gRNA.

**[0289]** The "guide sequence" also referred to as the "targeting sequence" of a gRNA can be modified so that the gRNA can target a CRISPR-Cas nuclease to any desired sequence of any desired target nucleic acid, with the exception that the protospacer adjacent motif (PAM) sequence can be taken into account. Thus, for example, a gRNA can have a targeting segment with a sequence (a guide sequence) that has complementarity with (e.g., can hybridize to) a sequence in a nucleic acid in a eukaryotic cell, e.g., a viral nucleic acid, a eukaryotic nucleic acid (e.g., a eukaryotic chromosome, chromosomal sequence, a eukaryotic RNA, etc.), and the like.

### ii. Double guide RNA (dgRNA)

**[0290]** In some embodiments, a gRNA includes two separate nucleic acid molecules: (1) a "targeter" and (2) an "activator." In these embodiments, the gRNA is also referred to as a "dual gRNA," a "double-molecule gRNA," a "two-molecule gRNA," a "dual gRNA," or a "dgRNA." In some embodiments, the targeter and the activator are covalently linked to one another (e.g., via intervening nucleotides) and the gRNA is referred to as a "single gRNA," a "Cas9 single gRNA," a "single-molecule Cas9 gRNA," a "one-molecule Cas9 gRNA," or simply "sgRNA."

**[0291]** In some embodiments, (1) the first molecule of the gRNA comprises a crRNA-like molecule ("CRISPR RNA," "targeter," "crRNA," or "crRNA repeat"), and (2) the second molecule of the gRNA comprises a tracrRNA-like molecule ("trans-acting CRISPR RNA," "activator," or "tracrRNA"). A dual gRNA can include any corresponding targeter and activator pair, or crRNA-like molecule and tracrRNA-like molecule pair.

**[0292]** The crRNA-like molecule (i.e., first molecule of a dgRNA or targeter) can comprise two portions – the first is a single-stranded targeting portion of the gRNA, and the second portion is a duplex-forming stretch of nucleotides ("duplex-forming portion") that forms one half of the dsRNA duplex with the tracrRNA-like molecule (i.e., second molecule of a dgRNA or activator) of the gRNA. Because the sequence of a targeting portion of the crRNA-like molecule (i.e., the portion that hybridizes with a target sequence of a target nucleic acid) is modified by a user to hybridize with a desired target nucleic acid (e.g., target locus on genomic DNA), the sequence of the targeting portion will often be a non-naturally occurring sequence.

**[0293]** In contrast, the duplex-forming portion of the crRNA-like molecule can include a naturally existing sequence. This is because the duplex-forming portion of the crRNA-like molecule (i.e., the first molecule of the dgRNA or targeter) hybridizes with the duplex-forming portion of a tracrRNA-like molecule (i.e., second molecule of the dgRNA or activator). The exact sequences of the duplex-forming portions of a given crRNA-like molecule and tracrRNA-like molecule pair are characteristic of the species in which the native RNA molecules are found. Examples of suitable cRNA and tracrRNA-related sequences are well known in the art. The sequence of the duplex-forming portion of a naturally existing crRNA can also be referred to as a crRNA repeat. Thus, the cRNA-like molecules of the present disclosure are distinguished from naturally occurring crRNAs, even though a cRNA-like molecule of the present disclosure (e.g., the duplex-forming segment) can include a naturally occurring sequence from a crRNA. However, the terms "cRNA-like molecule" and "targeter" encompass naturally occurring crRNAs.

**[0294]** As discussed in the preceding paragraph, a tracrRNA-like molecule (i.e., the second molecule of the dgRNA or activator) can form a duplex with a corresponding crRNA-like molecule. The tracrRNA-like molecule comprises a stretch of nucleotides (i.e., the duplex-forming portion) that hybridizes with the duplex-forming portion of an crRNA-like molecule. These hybridizing nucleotides of the tracrRNA-like molecule form one half of the dsRNA duplex of the

dgRNA with the crRNA-like molecule. Thus, each tracrRNA-like molecule of the present disclosure can bind to a corresponding crRNA-like molecule. The tracrRNA-like molecules of the present disclosure encompass naturally existing tracrRNAs and tracrRNAs with modifications (e.g., truncations, sequence variations, base modifications, backbone modifications, linkage modifications, etc.) where the activator retains at least one function of a tracrRNA (e.g., contributes to the dsRNA duplex to which Cas9 protein binds). In some embodiments, the tracrRNA-like molecule provides one or more stem loops that can interact with Cas9 protein.

### iii. Exemplary Guide Polynucleotides

**[0295]** In some embodiments, the guide polynucleotide is a CRISPR-Cas guide nucleic acid. In some embodiments, the CRISPR-Cas guide nucleic acid is a gRNA. The gRNA comprises (1) a protein-binding segment that binds to the CRISPR-Cas nuclease, and (2) a target-binding segment comprising a nucleotide sequence that is complementary to a target nucleotide sequence of a target site in a target cell (e.g., a locus in the genomic DNA of a target T cell, NK cell, B cell, monocyte, macrophage, dendritic cell, or HSC in a subject).

**[0296]** In some embodiments, the target cell is a T cell. Thus, in these embodiments, the gRNA target-binding segment comprises a nucleotide sequence that is complementary to a target locus in T cell genomic DNA. In some embodiments, the gRNA targets nucleotides that span the end of intron 1 and the beginning of exon 1 of TRAC. In some embodiments, the gRNA targets the CRISPR-Cas nuclease to cut the genomic DNA at 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 nucleotides before the beginning of exon 1. In some embodiments, the gRNA targets the CRISPR-Cas nuclease to cut the genomic DNA at 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 nucleotides after the beginning of exon 1. In some embodiments, the gRNA comprises the sequence of CAGGGTTCTGGATATCTGT (SEQ ID NO: 137) or TCAGGGTTCTGGATATCTGT (SEQ ID NO: 138) and targets the CRISPR-Cas nuclease to cut the genomic DNA at 1 nucleotide before the beginning of exon 1.

**[0297]** In some embodiments, the gRNA comprises a sequence having at least 70%, at least 80%, at least 85%, at least 90%, at least 95%, or 100% identity to SEQ ID NO: 137 or SEQ ID NO: 138. In some embodiments, the gRNA comprises the sequence of SEQ ID NO: 137 or 138. In some embodiments, the gRNA is under control of the U6 promoter. In some embodiments, an EDV, LNP, or AAV vector comprises a U6 promoter-TRAC-sgRNA scaffold comprises the following sequence:

GAGGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGATACAAGGCTGTTAGA
GAGATAATTAGAATTAATTTGACTGTAAACACAAAGATATTAGTACAAAATACGTG
ACGTAGAAAGTAATAATTTCTTGGGTAGTTTGCAGTTTTAAAATTATGTTTTAAAAT
GGACTATCATATGCTTACCGTAACTTGAAAGTATTTCGATTTCTTGGCTTTATATATC
TTGTGGAAAGGACGAAACACCGCAGGGTTCTGGATATCTGTGTTTTAGAGCTAGAA
ATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTC
GGTGCTTTTTT (SEQ ID NO: 139).

**[0298]** As noted previously, in some embodiments, the gRNA are delivered as part of a ribonucleoprotein (RNP) with the nuclease (e.g., Cas protein), e.g., with the EDV or LNP. In other embodiments, the gRNA is expressed from an expression cassette under the control of a promoter, e.g., from a polynucleotide in the AAV or the LNP.

### IV. Donor Template Polynucleotides

**[0299]** Disclosed herein are donor template polynucleotides for modifying the genome of a target cell. In many embodiments, a viral vector of the present disclosure, e.g., an AAV vector, comprises a donor template polynucleotide for delivery to a target cell. The donor template polynucleotide comprises a sequence for insertion (an "insertion sequence") into a target cell genome by a polynucleotide-guided nuclease, e.g., a CRISPR-Cas nuclease. The insertion sequence codes for a heterologous polypeptide that can be expressed by the target cell (i.e., a knock-in) after the insertion sequence is integrated into the target cell genome. In some embodiments, the donor template polynucleotide insertion sequence codes for a heterologous polypeptide comprising (a) an extracellular target-binding domain, (b) a transmembrane domain, (c) a hinge domain, and (d) an intracellular signaling domain. Examples of heterologous polypeptides are discussed below in detail. In some embodiments, integration of the coding sequence disrupts expression of the gene at the target site. In some embodiments, the donor template polynucleotide is a homology-directed repair template (HDRT) polynucleotide. In some embodiments, the donor template polynucleotide is a homology-independent targeted integration template (HITIT) polynucleotide. In some embodiments, the donor template polynucleotide is a homology-mediated end-joining template (HMEJT) polynucleotide.

**[0300]** In some embodiments, the integrated coding sequence is under control of endogenous regulatory sequences of the target gene, for example, endogenous promoter and/or enhancer sequences, to regulate expression of the heterologous polypeptide after the heterologous insertion sequence is integrated into the cell genome. In some embodiments, the insertion sequence includes

regulatory sequences, for example, promoter sequence and/or enhancer sequences, to regulate expression of the heterologous polypeptide after the insertion sequence is integrated into the cell genome. Exemplary insertion sequences and corresponding heterologous polypeptides are discussed below.

**[0301]** In some embodiments, the insertion sequence is integrated into or replaces an exon of a gene. In some embodiments, the insertion sequence is integrated into or replaces an intron of a gene. In some embodiments, the insertion sequence is integrated into or replaces an exon and an intron of a gene. In some embodiments, integration of the insertion sequence produces a fusion polypeptide that comprises (1) an endogenous polypeptide or a portion thereof of the target cell and (2) the heterologous polypeptide. In these embodiments, the endogenous polypeptide or portion thereof and the heterologous polypeptide are connected via a covalent linker. In some embodiments, the integrated coding sequence is under control of endogenous regulatory sequences of the target gene, for example, endogenous promoter and/or enhancer sequences, to regulate expression of the fusion polypeptide after the insertion sequence is integrated into the cell genome. In some embodiments, the insertion sequence includes regulatory sequences, for example, promoter sequence and/or enhancer sequences, to regulate expression of the fusion polypeptide after the insertion sequence is integrated into the cell genome.

**[0302]** In some embodiments, the donor template polynucleotide is a homology-directed repair template (HDRT) polynucleotide. The HDRT polynucleotide comprises one or two sequences with sufficient homology, i.e., two homology arms, to a sequence at the target site, e.g., at a target locus on genomic DNA, to support homology-directed repair between the HDRT polynucleotide and the target site which the homology arms have complementarity. The homology arms can contain at least 70%, at least 80%, at least 85%, at least 90%, at least 95%, at least 99%, or 100% homology with the nucleotide sequences flanking the target site, e.g., from 30 to 50 bases, from 15 to 30 bases, from 10 to 15 bases, or from 5 to 10 bases adjacent to or immediately flanking the target site. Each homology arm can comprise at least 50 nucleotides, at least 100 nucleotides, at least 150 nucleotides, at least 200 nucleotides, at least 250 nucleotides, at least 300 nucleotides, at least 350 nucleotides, at least 400 nucleotides, or at least 450 nucleotides (or any integral value between 10 and 450 nucleotides or more) of sequence homology with the sequence of a target site (e.g., a target locus on genomic DNA) so long as it can support homology-directed repair (e.g., for gene

insertion, gene deletion, gene disruption, or gene modification). The homology arms do not need to have 100% identity the sequences of the target site. Rather, each homology sequence may contain at least one or more single base changes, insertions, deletions, inversions, or rearrangements with respect to the sequence of a target site (e.g., a target locus on genomic DNA).

**[0303]** In some embodiments, the donor template polynucleotide is a homology-mediated end joining template (HMEJT) polynucleotide. The HMEJT polynucleotide comprises one or two sequences with sufficient homology, i.e., two homology arms, to a sequence at the target site, e.g., at a target locus on genomic DNA, to support homology-directed repair between the HMEJT polynucleotide and the target site which the homology arms have complementarity. The homology arms can contain at least 70%, at least 80%, at least 85%, at least 90%, at least 95%, at least 99%, or 100% homology with the nucleotide sequences flanking the target site, e.g., from 30 to 50 bases, from 15 to 30 bases, from 10 to 15 bases, or from 5 to 10 bases adjacent to or immediately flanking the target site. Each homology arm can comprise at least 50 nucleotides, at least 100 nucleotides, at least 150 nucleotides, at least 200 nucleotides, at least 250 nucleotides, at least 300 nucleotides, at least 350 nucleotides, at least 400 nucleotides, at least 450 nucleotides, at least 500 nucleotides, at least 550 nucleotides, at least 600 nucleotides, at least 650 nucleotides, at least 700 nucleotides, at least 750 nucleotides, at least 800 nucleotides, at least 850 nucleotides, at least 900 nucleotides, or at least 950 nucleotides (or any integral value between 10 and 950 nucleotides or more) of sequence homology with the sequence of a target site (e.g., a target locus on genomic DNA) so long as it can support homology-mediated repair (e.g., for gene insertion, gene deletion, gene disruption, or gene modification). The homology arms do not need to have 100% identity the sequences of the target site. Rather, each homology sequence may contain at least one or more single base changes, insertions, deletions, inversions, or rearrangements with respect to the sequence of a target site (e.g., a target locus on genomic DNA).

**[0304]** In some embodiments, the donor template polynucleotide is a homology-independent targeted integration template (HITIT) polynucleotide. The HITIT polynucleotide contains no sequences of homology with the target locus. In some embodiments, the HITIT polynucleotide also comprises a nuclease cleavage site (e.g., a CRISPR-Cas cleavage site) at either one or both ends of the HITIT polynucleotide insertion sequence.

**[0305]** In some embodiments, the target cell is an immune cell type, e.g., T cell, B cell, monocyte, macrophage, dendritic cell, and NK cell. In some embodiments, the target cell is a stem cell, e.g., an HSC. In general, the specific locus lies in a gene that is highly expressed in the target cell compared to the gene's expression in a non-target cell. For example, in embodiments where the target cell is a T cell, the target site is a gene with T-cell expression levels that are at least 50 times, at least 100 times, at least 250 times, at least 500 times, or at least 1,000 times its expression levels in a non-T cell. Thus, in some embodiments where the target cell is a T cell and the donor template polynucleotide is an HDRT polynucleotide or HMEJT polynucleotide, the HDRT polynueclotide or HMEJT polynucleotide comprises sequences with homology to portions of a gene that is expressed in a T cell many times, e.g., at least 50 times, at least 100 times, at least 250 times, at least 500 times, or at least 1,000 times over its expression in a non-target cell.

**[0306]** In some embodiments, the target cell is a T cell and the specific locus for sequence insertion is any gene in the T cell genome. In some embodiments, the gene is TRAC, TRBC, TRGC, TRDC, CD4, CD5, CD6, CD7, CD8a, CD8b, CD3e, CD247, CD27, CD28, IL-2R α, IL-2R beta, KLRC1, KLRF1, KLRG1, GZMA, GZMB, GZMH, GZMK, Zap-70, LCK, LAT, ITK, TCF7, LEF1, or FoxP3. Thus, in some embodiments, the HDRT polynucleotide or HMEJT polynucleotide comprises sequences with homology to portions of TRAC, TRBC, TRGC, TRDC, CD4, CD5, CD6, CD7, CD8a, CD8b, CD3e, CD247, CD27, CD28, IL-2R α, IL-2R beta, KLRC1, KLRF1, KLRG1, GZMA, GZMB, GZMH, GZMK, Zap-70, LCK, LAT, ITK, TCF7, LEF1, or FoxP3. In some embodiments, the gene is TRAC. In some embodiments, the target site is exon 1, exon 2, exon 3, exon 4, or any combination thereof, of TRAC. In some embodiments, the target site is an intron of TRAC. In some embodiments, the target site is exon 1 of TRAC and a first TRAC locus homology sequence is SEQ ID NO: 205 (shown below). In some embodiments, the target site is exon 1 of TRAC and a second TRAC locus homology sequence is SEQ ID NO: 206 (shown below).

**[0307]** In some embodiments, the target cell is a B cell and the specific locus for sequence insertion is any gene in the B cell genome. General methods for site-specific engineering B cells are discussed in Rogers, Geoffrey L, and Paula M Cannon. "Genome edited B cells: a new frontier in immune cell therapies." Molecular therapy: the journal of the American Society of Gene Therapy vol. 29,11 (2021): 3192-3204. doi:10.1016/j.ymthe.2021.09.019. In some embodiments,

the gene is CD19, CD20, CD22, CD138, BCMA, TACI, MS4A1, IGH, IGK, CD79A, or CD79B. Thus, in some embodiments, the HDRT polynucleotide or HMEJT polynucleotide comprises sequences with homology to portions of CD19, CD20, CD22, CD138, BCMA, TACI, MS4A1, IGH, IGK, CD79A, or CD79B.

**[0308]** In some embodiments, the target cell is an NK cell and the specific locus for sequence insertion is any gene in the NK cell genome. In some embodiments, the gene NCAM1, FCGR3A, NCR1, NCR2, KLRC1, NKG2D, NKG7, KIR2DL1, KIR2DL2, KIR2DL3, KIR2DL4, KIR3DL1 or KIR3DL1. Thus, in some embodiments, the HDRT polynucleotide or HMEJT polynucleotide comprises sequences with homology to portions of NCAM1, FCGR3A, NCR1, NCR2, KLRC1, NKG2D, NKG7, KIR2DL1, KIR2DL2, KIR2DL3, KIR2DL4, KIR3DL1 or KIR3DL1.

**[0309]** In some embodiments, the target cell is a monocyte or a macrophage and the specific locus for sequence insertion is any gene in the monocyte or macrophage T cell genome. In some embodiments, the gene is CD11b, CD11c, CD14, CD33, CD163, CLEC7A, C1QA, C1QB, C1QC, or MSR1. Thus, in some embodiments, the HDRT polynucleotide or HMEJT polynucleotide comprises sequences with homology to portions of CD11b, CD11c, CD14, CD33, CD163, CLEC7A, C1QA, C1QB, C1QC, or MSR1.

**[0310]** In some embodiments, the target cell is a dendritic cell and the specific locus for sequence insertion is any gene in the dendritic cell genome. General methods for engineering dendritic cells for cancer immunotherapy are discussed in Perez, Caleb R, and Michele De Palma. "Engineering dendritic cell vaccines to improve cancer immunotherapy." Nature communications vol. 10,1 5408. 27 Nov. 2019, doi:10.1038/s41467-019-13368-y. In some embodiments, the gene is CD1C, DCIR, CLEC10A, NDRG2, or TPM2. Thus, in some embodiments, the HDRT polynucleotide or HMEJT polynucleotide comprises sequences with homology to portions of CD1C, DCIR, CLEC10A, NDRG2, or TPM2.

**[0311]** In some embodiments, the target cell is an HSC and the specific locus for sequence insertion is any gene in the HSC genome. General methods for gene editing on HSCs are discussed in Ferrari, Samuele et al. "Gene Editing of Hematopoietic Stem Cells: Hopes and Hurdles Toward Clinical Translation." Frontiers in genome editing vol. 3 618378. 31 Mar. 2021, doi:10.3389/fgeed.2021.618378; and CRISPR/Cas9 β-globin gene targeting in human haematopoietic stem cells. In some embodiments, the gene is PTPRC, CD34, HBB, or RAG2.

Thus, in some embodiments, the HDRT polynucleotide or HMEJT polynucleotide comprises sequences with homology to portions of PTPRC, CD34, HBB, or RAG2.

**[0312]** In some embodiments, the donor template polynucleotide is covalently linked to a guide polynucleotide (e.g., gRNA). In these embodiments, the donor template polynucleotide can interact with and be bound by the polynucleotide-guided nuclease (e.g., CRISPR-Cas nuclease) via the guide polynucleotide (e.g., gRNA) to "shuttle" the donor template to the desired cellular location in proximity to the targeted nucleic acid to enhance gene modification efficiency. A guide polynucleotide can be covalently linked to the 5'-terminus or the 3'-terminus of the donor template polynucleotide.

**[0313]** In some embodiments, the donor template polynucleotide is provided to the cell (e.g., using a virus vector) as single-stranded DNA. In some embodiments, the donor template polynucleotide is provided to the cell as double-stranded DNA. The donor template polynucleotide can be introduced to the cell in linear or circular form. If introduced in linear form, the ends of the donor template polynucleotide may be protected (e.g., from exonucleolytic degradation) by any convenient method and such methods are known to those of skill in the art. For example, one or more dideoxynucleotide residues can be added to the 3' terminus of a linear molecule and/or self-complementary oligonucleotides can be ligated to one or both ends. See, for example, Chang et al. (1987) Proc. Natl. Acad Sci USA 84:4959-4963; Nehls et al. (1996) Science 272:886-889. Additional methods for protecting exogenous polynucleotides from degradation include, but are not limited to, addition of terminal amino group(s) and the use of modified internucleotide linkages such as, for example, phosphorothioates, phosphoramidates, and O-methyl ribose or deoxyribose residues. As an alternative to protecting the termini of a linear donor sequence, additional lengths of sequence may be included outside of the regions of homology that can be degraded without impacting recombination.

**[0314]** The donor template polynucleotides disclosed herein can be of any length, e.g. 50 nucleotides or more, 100 nucleotides or more, 200 nucleotides or more, 300 nucleotides or more, 400 nucleotides or more, 500 nucleotides or more, 1000 nucleotides or more, 5000 nucleotides or more, etc.

### a. Heterologous Polypeptides

**[0315]** Many heterologous polypeptides are suitable for expression in a target cell, and the coding sequence of many heterologous polypeptides are suitable for integration into the target cell genome. Thus, a donor template polynucleotide can comprise a nucleotide coding sequence (an "insertion sequence") for any suitable heterologous intracellular or cell surface polypeptide. In some embodiments, the insertion sequence of the donor template polynucleotide comprises a nucleotide coding sequence for a factor that modulates the immune system, a cytokine, a factor that modulates immune cell or stem cell function, a factor that promotes immune cell or stem cell survival, a factor that promotes immune cell or stem cell function, or an immune checkpoint inhibitor. A nucleotide coding sequence, particularly of a secreted protein or membrane-bound proteins, can include a nucleotide sequence encoding a signal peptide to target the extracellular domain of the cell receptor to the cell surface. The signal peptide can be endogenous to the protein encoded by the protein coding nucleotide sequence. The signal peptide can be heterologous to the protein encoded by the protein coding nucleotide sequence.

**[0316]** In embodiments where the target cell is a T cell, examples of heterologous polypeptides, without limitations, can include cell receptors, such as T cell receptors (TCRs), chimeric antigen receptors (CARs), synthetic Notch (synNotch) receptors, chimeric receptor synthetic intramembrane proteolysis receptors (SNIPRs), cytokines, and HLA-independent T cell receptors (HITs).

**[0317]** In embodiments where the target cell is a B cell, examples of heterologous polypeptides without limitations, can include antibodies and B cell receptors (BCRs).

**[0318]** In embodiments where the target cell is an NK cell, examples of heterologous polypeptides without limitations, can include CARs.

**[0319]** In embodiments where the target cell is a monocyte or a macrophage, examples of heterologous polypeptides without limitations, can include CARs.

**[0320]** In embodiments where the target cell is a dendritic cell, examples of heterologous polypeptides without limitations, can include chemokine receptors, CD40 ligand (CD40L; CD154), and chemokines.

**[0321]** In embodiments where the target cell is an HSC, examples of heterologous polypeptides without limitations, can include hemoglobin beta gene (HBB) sickle cell disease gene correction and RAG2 gene correction.

### i. Cell Receptors

**[0322]** In some embodiments, the donor template polynucleotide comprises a nucleotide coding sequence for a heterologous cell receptor polypeptide that comprises (1) an extracellular target-binding domain, (2) a transmembrane domain, (3) a hinge domain, and (4) an intracellular signaling domain.

**[0323]** The extracellular target-binding domain comprises a polypeptide that binds to a target polypeptide. In some embodiments, the extracellular antigen-binding domain of a comprises an scFv. The scFv can be derived from fusing the variable heavy and light regions ($V_H$ and $V_L$, respectively) of an antibody, or derived from a Fab or $F(ab)_2$. Other antibody-based recognition domains (cAb VHH (camelid antibody variable domains) and humanized versions, IgNAR VH (shark antibody variable domains) and humanized versions, sdAb VH (single domain antibody variable domains) and "camelized" antibody variable domains are suitable. In some embodiments, the antigen-binding domain is a nanobody. In some embodiments, the extracellular antigen-binding domain of a comprises a natural ligand of a target polypeptide, e.g., CD27 for binding to CD70, or a proligeration-inducing ligand (APRIL) for binding to B cell maturation antigen (BCMA) and transmembrane activator and CAML-interactor (TACI).

**[0324]** In some embodiments, the extracellular target-binding domain binds to a target polypeptide with a dissociation constant ($K_d$) of about $2\times10^{-7}$ M or less. In some embodiments, the $K_d$ is about $2\times10^{-7}$ M or less, about $1\times10^{-7}$ M or less, about $9\times10^{-8}$ M or less, about $1\times10^{-8}$ M or less, about $9\times10^{-9}$ M or less, about $5\times10^{-9}$ M or less, about $4\times10^{-9}$ M or less, about $3\times10^{-9}$ or less, about $2\times10^{-9}$ M or less, or about $1\times10^{-9}$ M or less. In certain non-limiting embodiments, the $K_d$ is about $3\times10^{-9}$ M or less. In certain non-limiting embodiments, the $K_d$ is from about $1\times10^{-9}$ M to about $3\times10^{-7}$ M. In certain non-limiting embodiments, the $K_d$ is from about $1.5\times10^{-9}$ M to about $3\times10^{-7}$ M. In certain non-limiting embodiments, the $K_d$ is from about $1.5\times10^{-9}$ M to about $2.7\times10^{-7}$ M. In certain non-limiting embodiments, the $K_d$ is from about $1\times10^{-4}$ M to about $1\times10^{-6}$ M. In certain non-limiting embodiments, the $K_d$ is from about $1\times10^{-13}$ M to about $1\times10^{-15}$ M.

**[0325]** In some embodiments, the extracellular target-binding domain binds to a target polypeptide associated with cancer, such as a cancer antigen or a tumor antigen, e.g., CD3, CD5, Claudin-6 (CLDN6), Claudin-18.2, CD19, MUC16, MUC1, CAIX, CEA, CD8, CD7, CD10, CD20, CD22, CD30, CLL1, CD33, CD34, CD38, CD41, CD44, CD49f, CD56, CD74, CD133, CD138, CD152, CD19, CD200, CD221, CD23 (igE receptor), CD28, CD4, CD40, CD51, CD52, CD80, EGP-2, EGP-40, EpCAM, erb-B2,3,4, FBP, Fetal acetylcholine receptor, folate receptor-a, GD2, GD3, HER-2, hTERT, IL-13R-a2, K-light chain, KDR, LeY, L1 cell adhesion molecule, MAGE-A1, Mesothelin, ERBB2, MAGEA3, p53, MART1, GP100, Proteinase3 (PR1), Tyrosinase, Survivin, hTERT, EphA2, NKG2D ligands, NY-ESO-1, oncofetal antigen (h5T4), PSCA, PSMA, ROR1, TAG-72, VEGF-A, VEGFR-1, VEGF-R2, WT-1, BCMA, CD123, CD44V6, NKCS1, EGF1R, EGFR-VIII, CD99, CD70, ADGRE2, CCR1, LILRB2, PRAME, CCR4, CD5, CD3, TRBC1, TRBC2, TIM-3, Integrin B7, ICAM-1, CD70, Tim3, CLEC12A, ERBB, Her2/neu, CA125, MUC-1, prostate-specific membrane antigen (PSMA), high molecular weight-melanoma associated antigen (HMW-MAA), 4-1BB, adenocarcinoma antigen, α-fetoprotein (AFP), BAFF, B-lymphoma cell, C242 antigen, carbonic anhydrase 9 (CA-IX), C-MET, CEA, FAP, ibronectin extra domain-B, folate receptor 1, GD2, GD3 ganglioside, glycoprotein 75, GPNMB, HER2/neu, HGF, human scatter factor receptorkinase, IGF-1 receptor,IGF-I, IgG1, L1-CAM, IL-13, IL-6, insulin-like growth factorI receptor,integrin α5β1, integrin αvβ3, MORAb-009, MS4A1, MUC1, mucin CanAg, N-glycolylneuraminic acid, NPC-1C, PDGF-R α, PDL192, phosphatidylserine, prostatic carcinoma cells, RANKL, RON, ROR1, SCH 900105, SDC1, SLAMF7, TAG-72, tenascin C, TGF beta 2, TGF-β, TRAIL-R1, TRAIL-R2, tumor antigen CTAA16.88, and vimentin. In some embodiments, the extracellular target-binding domain binds to a CD3 polypeptide.

**[0326]** In some embodiments, the extracellular target-binding domain comprises $V_H$ and $V_L$ amino acid sequences of cancer-associated antigen-binding antibodies; many examples are known in the art, as are the light chain and heavy chain CDRs of such antibodies. See, e.g., Ling et al. (2018) Frontiers Immunol. 9:469; International Patent Application Publication No. WO 2005/012493 and U.S. Patent Application Publication Nos. US 2019/0119375 and US 2013/0066055. The following are non-limiting examples of antibodies that bind cancer-associated antigens.

**[0327]** In some embodiments, the extracellular target-binding domain comprises an anti-CD19 antibody (e.g., an anti-CD19 scFv or an anti-CD19 nanobody). Anti-CD19 antibodies are known in the art; and the $V_H$ and $V_L$, or the $V_H$ and $V_L$ CDRs, of any anti-CD19 antibody can be included in a CAR. See e.g., International Patent Application Publication No. WO 2005/012493.

**[0328]** In some embodiments, an anti-CD19 antibody includes a $V_L$ CDR1 comprising the amino acid sequence KASQSVDYDGDSYLN (SEQ ID NO: 140); a $V_L$ CDR2 comprising the amino acid sequence DASNLVS (SEQ ID NO: 141); and a $V_L$ CDR3 comprising the amino acid sequence QQSTEDPWT (SEQ ID NO: 142). In some embodiments, an anti-CD19 antibody includes a $V_H$ CDR1 comprising the amino acid sequence SYWMN (SEQ ID NO: 143); a $V_H$ CDR2 comprising the amino acid sequence QIWPGDGDTNYNGKFKG (SEQ ID NO: 141); and a $V_H$ CDR3 comprising the amino acid sequence RETTTVGRYYYAMDY (SEQ ID NO: 145). In some embodiments, an anti-CD19 antibody includes a $V_L$ CDR1 comprising the amino acid sequence KASQSVDYDGDSYLN (SEQ ID NO: 146); a $V_L$ CDR2 comprising the amino acid sequence DASNLVS (SEQ ID NO: 147); a $V_L$ CDR3 comprising the amino acid sequence QQSTEDPWT (SEQ ID NO: 148); a $V_H$ CDR1 comprising the amino acid sequence SYWMN (SEQ ID NO: 149); a $V_H$ CDR2 comprising the amino acid sequence QIWPGDGDTNYNGKFKG (SEQ ID NO: 150); and a $V_H$ CDR3 comprising the amino acid sequence RETTTVGRYYYAMDY (SEQ ID NO: 151).

**[0329]** In some embodiments, an anti-CD19 antibody is a scFv. For example, in some embodiments, an anti-CD19 scFv comprises an amino acid sequence having at least 90%, at least 95%, at least 98%, at least 99%, or 100%, amino acid sequence identity to the following amino acid sequence as set forth in SEQ ID NO: 152 shown below.

DIQLTQSPASLAVSLGQRATISCKASQSVDYDGDSYLNWYQQIPGQPPKLLIYDASNLVS GIPPRFSGSGSGTDFTLNIHPVEKVDAATYHCQQSTEDPWTFGGGTKLEIKGGGGSGGG GSGGGGSQVQLQQSGAELVRPGSSVKISCKASGYAFSSYWMNWVKQRPGQGLEWIGQI WPGDGDTNYNGKFKGKATLTADESSSTAYMQLSSLASEDSAVYFCARRETTTVGRYYY AMDYWGQGTTVTVS (SEQ ID NO: 152).

**[0330]** The transmembrane domain links the extracellular target-binding domain and the intracellular signaling domain, and anchors the heterologous polypeptide to the plasma membrane of the host cell that is modified to express the heterologous polypeptide (e.g., the plasma membrane

of a human T cell, B cell, NK cell, monocyte, macrophage, dendritic cell, or HSC). Any transmembrane domain suitable for use in a cell receptor construct may be employed. A transmembrane domain incorporated into a cell receptor construct may be derived either from a natural, synthetic, semi-synthetic, or recombinant source. Such transmembrane domains, include, but are not limited to, all or part of the transmembrane domain of the α, beta or ζ chain of the T-cell receptor, CD28, CD27, CD3 epsilon, CD45, CD4, CD5, CD8, CD9, CD16, CD22, CD33, CD37, CD64, CD80, CD86, CD134, CD137, CD154. In some embodiments, a transmembrane domain may include at least the transmembrane region(s) of, e.g., KIRDS2, OX40, CD2, CD27, LFA-1 (CD 11a, CD18), ICOS (CD278), 4-1BB (CD137), GITR, CD40, BAFFR, HVEM (LIGHTR), SLAMF7, NKp80 (KLRF1), NKp44, NKp30, NKp46, CD160, CD19, IL2R beta, IL2R gamma, IL7R a, ITGAl, VLAl, CD49a, ITGA4, IA4, CD49D, ITGA6, VLA-6, CD49f, ITGAD, CDl ld, ITGAE, CD103, ITGAL, CDl la, LFA-1, ITGAM, CDl lb, ITGAX, CDl lc, ITGB 1, CD29, ITGB2, CD 18, LFA-1, ITGB7, TNFR2, DNAM1 (CD226), SLAMF4 (CD244, 2B4), CD84, CD96 (Tactile), CEACAM1, CRT AM, Ly9 (CD229), CD160 (BY55), PSGL1, CD100, (SEMA4D), SLAMF6 (NTB-A, Lyl08), SLAM (SLAMF1, CD150, IPO-3), BLAME, (SLAMF8), SELPLG (CD162), LTBR, PAG/Cbp, NKG2D, or NKG2C. In some embodiments, the transmembrane domain is derived from the CD28 polypeptide.

**[0331]** The hinge domain links the extracellular target-binding domain and the transmembrane domain for positioning the extracellular target-binding domain. In some embodiments, the cell receptor may contain one or more hinge domains that link the extracellular target-binding domain and the transmembrane domain for positioning the extracellular target-binding domain. Such a hinge domain may be derived either from a natural, synthetic, semi-synthetic, or recombinant source. The hinge domain can include the amino acid sequence of a naturally occurring immunoglobulin hinge region, e.g., a naturally occurring human immunoglobulin hinge region, or an altered immunoglobulin hinge region. Illustrative hinge domains suitable for use in a cell receptors include the hinge region derived from the extracellular regions of type 1 membrane proteins such as CD8 α, CD4, CD28, PD1, CD152, CD166 and CD7, which may be wildtype hinge regions from these molecules or may be altered. In some embodiments, the hinge domain is based on the hinge region of a human immunoglobulin IgG1 or IgG4. In some embodiments, the hinge region includes the IgG1 or IgG4's CH2 region, which may comprise one or more mutations. In some embodiments, the mutation is L235E, N297Q, or both L235E and N297Q (which is known

as the EQ mutation in the IgG4 hinge region). In some embodiments, the hinge domain is derived from the CD28 polypeptide.

**[0332]** A cell receptor can include one or more intracellular signaling domains, also referred to herein as co-stimulatory domains, activation domains, or cytoplasmic domains that activate or otherwise modulate an immune cell. The intracellular signaling domain is generally responsible for activation of at least one of the normal effector functions of the immune cell in which the cell receptor has been introduced. For example, a "first-generation chimeric antigen receptor" ("CAR") generally has a CD3 zeta (CD3ζ) signaling domain. Additional costimulatory intracellular domains may also be introduced (e.g., second and third generation CARS) and further domains including homing and suicide domains may be included in CAR constructs.

**[0333]** In some embodiments, an intracellular signaling domain is used that increases immune cell (e.g., T cell or macrophage) cytokine production. In some embodiments, an intracellular signaling domain is used that facilitates immune cell (e.g., T cell, B cell, NK cell, monocyte, macrophage, or dendritic cell) replication. In some embodiments, an intracellular signaling domain is used that prevents immune cell (e.g., T cell) exhaustion. In some embodiments, an intracellular signaling domain is used that increases immune cell (e.g., T cell, B cell, NK cell, or dendritic cell) antitumor activity. In some embodiments, an intracellular signaling domain is used that enhances survival of immune cells (e.g., T cells, B cells, NK cells, monocytes, macrophages, or dendritic cells) (e.g., post-infusion into patients). Examples of intracellular signaling domains for use in a cell receptor include the cytoplasmic sequences of a cell receptor (e.g., T cell receptor (TCR) and B cell receptor (BCR)) and co-receptors that act in concert to initiate signal transduction following antigen receptor engagement, as well as any derivative or variant of these sequences and any recombinant sequence that has the same functional capability.

**[0334]** A primary intracellular signaling domain regulates primary activation of the TCR complex either in a stimulatory way, or in an inhibitory way. Primary intracellular signaling domains that act in a stimulatory manner may contain signaling motifs which are known as immunoreceptor tyrosine-based activation motifs or ITAMs. Examples of ITAM containing primary intracellular signaling domains include those of CD3ζ, common FcR gamma, Fc gamma Rlla, FcR beta (Fc Epsilon Rib), CD3 gamma, CD3 delta, CD3 epsilon, CD79a, CD79b, DAP10, and DAP12. Nonlimiting examples of ITAM sequences are shown below in Table 6.

| Table 6 – Exemplary ITAM Sequences | | |
|---|---|---|
| **SEQ ID NO** | **Description** | **Sequence** |
| 153 | Exemplary ITAM1 | QNQLYNELNLGRREEYDVLDKR |
| 154 | Exemplary ITAM1 | QNQLFNELNLGRREEFDVLDKR |
| 155 | Exemplary ITAM2 | QEGLYNELQKDKMAEAYSEIGMK |
| 156 | Exemplary ITAM2 | QEGLFNELQKDKMAEAFSEIGMK |
| 157 | Exemplary ITAM3 | HDGLYQGLSTATKDTYDALHMQ |
| 158 | Exemplary ITAM3 | HDGLFQGLSTATKDTFDALHMQ |

**[0335]** An intracellular signaling domain of a cell receptor can comprise a primary intracellular signaling domain only, or may comprise additional desired intracellular signaling domain(s). For example, the intracellular signaling domain of a cell receptor can comprise a CD3ζ chain portion and a costimulatory signaling domain. The costimulatory signaling domain refers to a portion of the cell receptor comprising the intracellular domain of a costimulatory molecule. A costimulatory molecule is a cell surface molecule other than an antigen receptor or its ligands that is required for an efficient response of lymphocytes to an antigen. Examples of such molecules include CD27, CD28, 4-1BB (CD137), OX40, CD30, CD40, PD-1, ICOS, lymphocyte function-associated antigen-1 (LFA-1), CD2, CD7, LIGHT, NKG2C, B7-H3, and a ligand that binds to CD83, and the like. For example, CD27 costimulation has been demonstrated to enhance expansion, effector function, and survival of human CART cells in vitro and augments human T cell persistence and antitumor activity in vivo (Song et al. Blood. 2012; 119(3):696-706). Further examples of such costimulatory molecules include CDS, ICAM-1, GITR, BAFFR, HVEM (LIGHTR), SLAMF7, NKp80 (KLRF1), NKp44, NKp30, NKp46, CD 160, CD 19, CD4, CD8α, CD8β, IL2Rbeta, IL2R gamma, IL7R α, ITGA4, VLA1, CD49a, ITGA4, IA4, CD49D, ITGA6, VLA-6, CD49f, ITGAD, CDl ld, ITGAE, CD103, ITGAL, CDl la, LFA-1, ITGAM, CDl lb, ITGAX, CDl lc, ITGB 1, CD29, ITGB2, CD18, LFA-1, ITGB7, TNFR2, TRANCE/RANKL, DNAM1 (CD226), SLAMF4

(CD244, 2B4), CD84, CD96 (Tactile), NKG2D, CEACAMl, CRTAM, Ly9 (CD229), CD160 (BY55), PSGL1, CD100 (SEMA4D), CD69, SLAMF6 (NTB-A, Lyl08), SLAM, (SLAMFl, CD150, IPO-3), BLAME (SLAMF8), SELPLG (CD162), LTBR, LAT, GADS, SLP-76, PAG/Cbp, and CD19a.

**[0336]** In some embodiments, the intracellular signaling domain of the cell receptor comprises a modified CD3ζ polypeptide comprising or consisting essentially of or consisting of an ITAM1 variant comprising one or more loss-of-function mutations, an ITAM2 variant comprising one or more loss-of-function mutations, an ITAM3 variant comprising one or more loss-of-function mutations, or a combination thereof. In some embodiments, the loss-of-function mutation comprises a Tyr to Phe mutation.

**[0337]** In some embodiments, the intracellular signaling domain of the cell receptor comprises a modified CD3ζ polypeptide comprising an ITAM2 variant comprising one or more (e.g., two) loss-of-function mutations and an ITAM3 variant comprising one or more (e.g., two) loss-of-function mutations. In some embodiments, the intracellular signaling domain of the cell receptor comprises a modified CD3ζ polypeptide comprising a native ITAM1, an ITAM2 variant comprising two loss-of-function mutations and an ITAM3 variant comprising two loss-of-function mutations. In some embodiments, the intracellular signaling domain of the cell receptor comprises a modified CD3ζ polypeptide comprising a native ITAM1 having the amino acid sequence set forth in SEQ ID NO: 153 , an ITAM2 variant having the amino acid sequence set forth in SEQ ID NO: 156 and an ITAM3 variant having the amino acid sequence set forth in SEQ ID NO: 158 (e.g., a construct designated as "1XX").

**[0338]** In some embodiments, the intracellular signaling domain of the cell receptor comprises a modified CD3ζ polypeptide comprising an ITAM1 variant comprising one or more (e.g., two) loss-of-function mutations and an ITAM3 variant comprising one or more (e.g., two) loss-of-function mutations. In some embodiments, the intracellular signaling domain of the cell receptor comprises a modified CD3ζ polypeptide comprising an ITAM1 variant comprising two loss-of-function mutations, a native ITAM2, and an ITAM3 variant comprising two loss-of-function mutations. In some embodiments, the intracellular signaling domain of the cell receptor comprises a modified CD3ζ polypeptide comprising an ITAM1 variant having the amino acid sequence set forth in SEQ ID NO: 154, a native ITAM2 having the amino acid sequence set forth in SEQ ID

NO: 155 and an ITAM3 variant having the amino acid sequence set forth in SEQ ID NO: 158 (e.g., a construct designated as "X2X").

**[0339]** In some embodiments, the intracellular signaling domain of the cell receptor comprises a modified CD3ζ polypeptide comprising an ITAM1 variant comprising one or more (e.g., two) loss-of-function mutations and an ITAM2 variant comprising one or more (e.g., two) loss-of-function mutations. In some embodiments, the intracellular signaling domain of the cell receptor comprises a modified CD3ζ polypeptide comprising an ITAM1 variant comprising two loss-of-function mutations, an ITAM2 variant comprising two loss-of-function mutations, and a native ITAM3. In some embodiments, the intracellular signaling domain of the cell receptor comprises a modified CD3ζ polypeptide comprising an ITAM1 variant having the amino acid sequence set forth in SEQ ID NO: 154 , an ITAM2 variant having the amino acid sequence set forth in SEQ ID NO: 156 and a native ITAM3 having the amino acid sequence set forth in SEQ ID NO: 157 (e.g., a construct designated as "XX3").

**[0340]** In some embodiments, the intracellular signaling domain of the cell receptor comprises a modified CD3ζ polypeptide comprising a native ITAM1, a native ITAM2, and an ITAM3 variant comprising one or more (e.g., two) loss-of-function mutations. In some embodiments, the intracellular signaling domain of the cell receptor comprises a modified CD3ζ polypeptide comprising a native ITAM1, a native ITAM2, and an ITAM1 variant comprising two loss-of-function mutations. In some embodiments, the intracellular signaling domain of the cell receptor comprises a modified CD3ζ polypeptide comprising a native ITAM1 having the amino acid sequence set forth in SEQ ID NO: 153, a native ITAM2 having the amino acid sequence set forth in SEQ ID NO: 155 and an ITAM3 variant having the amino acid sequence set forth in SEQ ID NO: 158 (e.g., a construct designated as "12X").

**[0341]** In some embodiments, the intracellular signaling domain of the cell receptor comprises a modified CD3ζ polypeptide comprising a native ITAM1, an ITAM2 variant comprising one or more (e.g., two) loss-of-function mutations, and a native ITAM3. In some embodiments, the intracellular signaling domain of the cell receptor comprises a modified CD3ζ polypeptide comprising a native ITAM1, an ITAM2 variant comprising two loss-of-function mutations, and a native ITAM3. In some embodiments, the intracellular signaling domain of the cell receptor comprises a modified CD3ζ polypeptide comprising a native ITAM1 having the amino acid

sequence set forth in SEQ ID NO: 153, an ITAM2 variant having the amino acid sequence set forth in SEQ ID NO: 156 and a native ITAM3 variant having the amino acid sequence set forth in SEQ ID NO: 157 (e.g., a construct designated as "1X3").

**[0342]** In some embodiments, the intracellular signaling domain of the cell receptor comprises a modified CD3ζ polypeptide comprising a deletion of one or two ITAMs. In some embodiments, the modified CD3ζ polypeptide comprises a deletion of ITAM1 and ITAM2, e.g., the modified CD3ζ polypeptide comprises a native ITAM3 or a ITAM3 variant, and does not comprise an ITAM1 or an ITAM2. In some embodiments, the modified CD3ζ polypeptide comprises a native ITAM3 having the amino acid sequence set forth in SEQ ID NO: 157, and does not comprise an ITAM1 (native or modified), or an ITAM2 (native or modified) (e.g., a construct designated as D12).

**[0343]** In some embodiments, the modified CD3ζ polypeptide comprises a deletion of ITAM2 and ITAM3, e.g., the modified CD3ζ polypeptide comprises a native ITAM1 or a ITAM1 variant, and does not comprise an ITAM2 or an ITAM3. In some embodiments, the modified CD3ζ polypeptide comprises a native ITAM1 having the amino acid sequence set forth in SEQ ID NO: 153, and does not comprise an ITAM2 (native or modified), or an ITAM3 (native or modified) (e.g., a construct designated as D23).

**[0344]** In some embodiments, the modified CD3ζ polypeptide comprises a deletion of ITAM1 and ITAM3, e.g., the modified CD3ζ polypeptide comprises a native ITAM2 or a ITAM2 variant, and does not comprise an ITAM1 or an ITAM3. In some embodiments, the modified CD3ζ polypeptide comprises a native ITAM2 having the amino acid sequence set forth in SEQ ID NO: 155, and does not comprise an ITAM1 (native or modified), or an ITAM3 (native or modified) (e.g., a construct designated as D13).

**[0345]** In some embodiments, the modified CD3ζ polypeptide comprises a deletion of ITAM1, e.g., the modified CD3ζ polypeptide comprises a native ITAM2 or an ITAM2 variant, and a native ITAM3 or an ITAM3 variant, and does not comprise an ITAM1 (native or modified).

**[0346]** In some embodiments, the modified CD3ζ polypeptide comprises a deletion of ITAM2, e.g., the modified CD3ζ polypeptide comprises a native ITAM1 or an ITAM1 variant, and a native ITAM3 or an ITAM3 variant, and does not comprise an ITAM2 (native or modified).

**[0347]** In some embodiments, the modified CD3ζ polypeptide comprises a deletion of ITAM3, e.g., the modified CD3ζ polypeptide comprises a native ITAM1 or an ITAM1 variant, and a native ITAM2 or an ITAM2 variant, and does not comprise an ITAM3 (native or modified).

**[0348]** In some embodiments, the intracellular signaling domain of the cell receptor comprises a modified CD3ζ polypeptide comprising an ITAM1 variant comprising one or more (e.g., two) loss-of-function mutations. In some embodiments, the intracellular signaling domain of the CAR comprises a modified CD3ζ polypeptide comprising an ITAM1 variant comprising two loss-of-function mutations, a native ITAM2, and a native ITAM3. In some embodiments, the intracellular signaling domain of the CAR comprises a modified CD3ζ polypeptide comprising an ITAM1 variant having the amino acid sequence set forth in SEQ ID NO: 155, a native ITAM2 having the amino acid sequence set forth in SEQ ID NO: 156 and a native ITAM3 having the amino acid sequence set forth in SEQ ID NO: 157 (e.g., a construct designated as "X23").

**[0349]** The heterologous polypeptide may also comprise a signal peptide to target the extracellular domain of the cell receptor to the cell surface. Thus, the donor template polynucleotide may also comprise a nucleotide coding sequence for a signal peptide.

### 1. T Cell Receptors (TCRs) and Variants

**[0350]** Insertion of a heterologous nucleotide coding sequence into the TCR locus means that the expression of the heterologous protein will be controlled by the endogenous TCR promoter and in some embodiments will be expressed as part of a larger fusion protein with a TCR polypeptide that is subsequently cleaved to form separate TCR and heterologous polypeptides. As noted earlier, integration of a nucleotide sequence can produce a heterologous cell receptor, or add to a TCR locus to provide a TCR variant (e.g., to provide a different affinity, for example, but not limited to, for a cancer antigen) to the T cell. In some embodiments, the heterologous nucleotide sequence is inserted in an exon of the TRAC, TRBC, TRGC, or TRDC gene. In some embodiments, the heterologous nucleotide sequence is inserted in an intron of the TRAC, TRBC, TRGC, or TRDC gene. In some embodiments, the heterologous nucleotide sequence is inserted in exon 1 of the TRAC gene. In some embodiments, the heterologous nucleotide sequence is inserted in exon 1 of the TRBC gene. Upon insertion of the heterologous nucleotide sequence into the TCR locus of a cell, the heterologous nucleotide sequence is under the control of an endogenous TCR promoter, for example a TRAC promoter or a TRBC promoter. Thus, the heterologous nucleotide

sequence provided herein encode a TCR variant that is co-expressed with the heterologous polypeptide. Once the heterologous nucleotide sequence is incorporated into the genome of the T cell by HDR, and under the control of the endogenous promoter, the T cells can transcribe the target gene, including the inserted heterologous nucleotide sequence, into a single mRNA sequence encoding a fusion polypeptide that is then processed into separate heterologous polypeptides (e.g., for example by cleavage of one or more peptide sequences linking the polypeptides). In some embodiments, insertion of a heterologous nucleotide sequence of the present disclosure will produce a T cell with a fusion polypeptide with the function and/or specificity of the TCR combined with the function and/or specificity of the heterologous polypeptide.

**[0351]** As used throughout, the term "endogenous TCR subunit" is the TCR subunit, for example, TCR-α or TCR-β that is endogenously expressed by the cell that the heterologous nucleotide sequence is introduced into. As set forth above, integration of a heterologous nucleotide sequence into a TCR locus produces a modified nucleotide sequence that encodes multiple amino acid sequences that are expressed as a multicistronic sequence. This multicistronic sequence is translated into a single polypeptide chain that is processed, i.e., self-cleaved, to produce two, three, four, five, six, seven, or more separate polypeptide chains. For example one or more TCR subunits and the heterologous polypeptide encoded by the inserted nucleotide sequence may be produced. In another example, integration of a heterologous nucleotide sequence into a target locus produces a synthetic cell receptor, e.g., a chimeric antigen receptor (CAR) that is encoded by the inserted nucleotide sequence.

**[0352]** Examples of self-cleaving peptides include, but are not limited to, self-cleaving viral 2A peptides, for example, a porcine teschovirus-1 (P2A) peptide, a Thosea asigna virus 2A-like peptide (T2A), an equine rhinitis A virus (E2A) peptide, or a foot-and-mouth disease virus (F2A) peptide. Self-cleaving 2A peptides allow expression of multiple gene products from a single construct. (See, for example, Chng et al. "Cleavage efficient 2A peptides for high level monoclonal antibody expression in CHO cells," *MAbs* 7(2): 403-412 (2015)). In some embodiments, the nucleic acid construct comprises two or more self-cleaving peptides. In some embodiments, the two or more self-cleaving peptides are all the same. In other embodiments, at least one of the two or more self-cleaving peptides is different.

**[0353]** In some embodiments, when targeting to a human TCR locus, the size of the insertion sequence encoding the N-terminal portion of the endogenous TCR subunit will depend on the number of nucleotides in the endogenous TCR gene locus, e.g., TRAC or TRBC nucleic acid sequence between the start of TRAC exon 1 or TRBC exon 1 and the targeted insertion site. For example, if the number of nucleotides between the start of TRAC exon 1 and the insertion site is less than or greater than 25 nucleotides, an insertion sequence of less than or greater than 25 nucleotides encoding the N-terminal portion of the endogenous TCR-α subunit can be in the donor template polynucleotide.

**[0354]** In some embodiments, when targeting to a human TCR locus and the donor template polynucleotide is an HDRT polynucleotide or HMEJT polynucleotide, the HDRT polynucleotide or HMEJT polynucleotide comprises flanking homology arm sequences having homology to a human TCR locus. In some embodiments, the length of each homology arm sequence is at least about 50, 100, 150, 200, 250, 300, 350, 400 or 450 nucleotides. In some embodiments, a nucleotide sequence that is homologous to a genomic sequence is at least 70%, 80%, 85% 90%, 95%, 99% or 100% complementary to the genomic sequence. In some embodiments, one or both homology arm sequences optionally comprises a mismatched nucleotide sequence compared to a homologous sequence in the genomic sequence in the TCR locus flanking the insertion site in the TCR locus.

**[0355]** In some embodiments, the donor template polynucleotide comprises a coding sequence for a portion of a TCR, such that integration of that coding sequence into the target locus produces a TCR variant. An endogenous TCR of a T cell may be modified to alter the specificity of the TCR for targeting cell-surface antigens. TCR modification can be achieved by targeting the donor template polynucleotide to a locus in a TCR gene, e.g., TRAC, TRBC, TRGC, and TRDC, such that integration of a heterologous coding sequence at the chosen locus produces a TCR variant comprising (1) a portion of an endogenous TCR and (2) the heterologous polypeptide. In these embodiments, the endogenous TCR portion and the heterologous polypeptide are covalently linked and are expressed as a fusion polypeptide. In some embodiments, the fusion polypeptide is processed into separate heterologous polypeptides (e.g., for example by cleavage of one or more peptide sequences linking the polypeptides). For example, the TRAC locus of a T cell can be edited such that the extracellular domain of the TRAC polypeptide is replaced with a heterologous extracellular target-binding domain. In this way, the specificity, sensitivity, and/or function of the

extracellular target-binding domain of the TCR is modified. Examples of modified TCR fusion polypeptides include HLA-independent T cell receptors (HITs). Thus, in some embodiments, the donor template polynucleotide comprises a coding sequence for a portion of a HIT. In some embodiments, the target site is an intron of TRAC. In some embodiments, the target site is exon 1 of TRAC and a first TRAC locus homology sequence is SEQ ID NO: 205. In some embodiments, the target site is exon 1 of TRAC and a second TRAC locus homology sequence is SEQ ID NO: 206.

**[0356]** In some embodiments, the donor template polynucleotide comprises a coding sequence for a chimeric antigen receptor (CAR). In some embodiments, the heterologous nucleotide coding sequence for a CAR is targeted to a TCR locus. Exemplary CAR structures and methods for making them are known in the art; see e.g., International Patent Application Publication Nos. WO 2009/091826, WO 2015/142675, WO 2014/055657, and WO 2015/090229; U.S. Patent Application Publication Nos. 2020/0317777, 2013/0287748, 2016/0185861, and 2019/0000880; and U.S. Patent No. 9,587,020. In some embodiments, the CAR is a single polypeptide chain. In some embodiments, the CAR comprises two polypeptide chains. Generally, any CAR structure known to those skilled in the art can be used. Exemplary CARs are discussed below.

### 2. Exemplary Chimeric Antigen Receptors (CARs)

**[0357]** In some embodiments, the heterologous cell receptor is a chimeric antigen receptor (CAR). Thus, in some embodiments, the donor template polynucleotide comprises a coding sequence for a CAR. Non-limiting examples of CARs are discussed in U.S. Patent Application Publication No. 2020/0317777.

**[0358]** In some embodiments, the heterologous cell receptor is a CAR comprising an extracellular antigen-binding domain that binds to a CD19 polypeptide (e.g., a human CD19 polypeptide), a transmembrane domain and a hinge/spacer region derived from a CD28 polypeptide, an intracellular signaling domain comprising a modified CD3ζ polypeptide (e.g., a modified human CD3ζ polypeptide) comprising a native ITAM1, a native ITAM2, a native ITAM3, a native BRS1, a native BRS2, and a native BRS3, and a co-stimulatory signaling region comprising a CD28 polypeptide (e.g., a human CD28 polypeptide). In some embodiments, the CAR is designated as "1928z WT." In some embodiments, the CAR (e.g., 1928z WT) comprises an amino acid sequence that is at least about 80%, at least about 85%, at least about 90%, at least

about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or at least about 100% homologous or identical to the amino acid sequence set forth in SEQ ID NO: 159, which is provided in Table 7 below. SEQ ID NO: 159 includes a CD8 leader sequence at amino acids 1 to 18, and is able to bind to CD19 (e.g., human CD19).

**[0359]** In some embodiments, the heterologous cell receptor is a CAR comprising an extracellular antigen-binding domain that binds to a CD19 polypeptide (e.g., a human CD19 polypeptide), a transmembrane domain and a hinge/spacer region derived from a CD28 polypeptide, an intracellular signaling domain comprising a modified CD3ζ polypeptide (e.g., a modified human CD3ζ polypeptide) comprising a native ITAM1, a native BRS1, a native BRS2, a native BRS3, an ITAM2 variant having two loss-of-function mutations, and an ITAM3 variant having two loss-of-function mutations, and a co-stimulatory signaling region comprising a CD28 polypeptide (e.g., a human CD28 polypeptide). In some embodiments, the CAR is designated as "1928z-1XX". In some embodiments, the CAR (e.g., 1928z-1XX) comprises an amino acid sequence that is at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or at least about 100% homologous or identical to the amino acid sequence set forth in SEQ ID NO: 160, which is provided in Table 7 below. SEQ ID NO: 160 includes a CD8 leader sequence at amino acids 1 to 18, and is able to bind to CD19 (e.g., human CD19).

**[0360]** In some embodiments, the heterologous cell receptor is a CAR comprising an extracellular antigen-binding domain that binds to a CD19 polypeptide (e.g., a human CD19 polypeptide), a transmembrane domain and a hinge/spacer region derived from a CD28 polypeptide, an intracellular signaling domain comprising a modified CD3ζ polypeptide (e.g., a modified human CD3ζ polypeptide), and a co-stimulatory signaling region comprising a CD28 polypeptide (e.g., a human CD28 polypeptide), wherein the modified CD3ζ polypeptide comprises a native ITAM3 and does not comprise an ITAM1 (native or modified), an ITAM2 (native or modified), a BRS1 (native or modified), a BRS2 (native or modified), or a BRS3 (native or modified). In someembodiments, the CAR is designated as "D12". In someembodiments, the CAR (e.g., D12) comprises an amino acid sequence that is at least about 85%, about 90%, about 95%, about 96%, about 97%, about 98%, about 99% or about 100% homologous to the amino acid sequence set forth in SEQ ID NO: 161, which is provided in Table 7 below. SEQ ID NO: 161

includes a CD8 leader sequence at amino acids 1 to 18, and is able to bind to CD19 (e.g., human CD19).

**[0361]** In some embodiments, the heterologous cell receptor is a CAR comprising an extracellular antigen-binding domain that binds to a CD19 polypeptide (e.g., human CD19 polypeptide), a transmembrane domain and a hinge/spacer region derived from a CD28 polypeptide, an intracellular signaling domain comprising a modified CD3ζ polypeptide (e.g., a modified human CD3ζ polypeptide) comprising ITAM1, BRS1 and a deletion of ITAM2, ITAM3, BRS2 and BRS3, and a co-stimulatory signaling region comprising a CD28 polypeptide (e.g., a human CD28 polypeptide), wherein the modified CD3ζ polypeptide comprises a native ITAM1 and a native BRS1, and does not comprise an ITAM2 (native or modified), an ITAM3 (native or modified), a BRS2 (native or modified), or a BRS3 (native or modified). In some embodiments, the CAR is designated as "D23". In some embodiments, the CAR (e.g., D23) comprises an amino acid sequence that is at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or at least about 100% homologous or identical to the amino acid sequence set forth in SEQ ID NO: 162, which is provided in Table 7 below. SEQ ID NO: 162 includes a CD8 sequence at amino acids 1 to 18, and is able to bind to CD19 (e.g., human CD19).

**[0362]** In some embodiments, the heterologous cell receptor is a CAR comprising an extracellular antigen-binding domain that binds to a CD19 polypeptide (e.g., a human CD19 polypeptide), a transmembrane domain and a hinge/spacer region derived from a CD28 polypeptide, an intracellular signaling domain comprising a modified CD3ζ polypeptide (e.g., a modified human CD3ζ polypeptide) comprising a native ITAM3, a native BRS1, a native BRS2, a native BRS3, an ITAM1 variant having two loss-of-function mutations, and an ITAM2 variant having two loss-of-function mutations, and a co-stimulatory signaling region comprising a CD28 polypeptide (e.g., a human CD28 polypeptide). In some embodiments, the CAR is designated as "XX3". In some embodiments, the CAR (e.g., XX3) comprises an amino acid sequence that is at least about 85%, about 90%, about 95%, about 96%, about 97%, about 98%, about 99% or about 100% homologous to the amino acid sequence set forth in SEQ ID NO: 163, which is provided in Table 7 below. SEQ ID NO: 163 includes a CD8 leader sequence at amino acids 1 to 18, and is able to bind to CD19 (e.g., human CD19).

**[0363]** In some embodiments, the heterologous cell receptor is a CAR comprising an extracellular antigen-binding domain that binds to a CD19 polypeptide (e.g., a human CD19 polypeptide), a transmembrane domain and a hinge/spacer region derived from a CD28 polypeptide, an intracellular signaling domain comprising a modified CD3ζ polypeptide (e.g., a modified human CD3ζ polypeptide) comprising a native ITAM2, a native ITAM3, a native BRS1, a native BRS2, a native BRS3, and an ITAM1 variant having two loss-of-function mutations, and a co-stimulatory signaling region comprising a CD28 polypeptide (e.g., a human CD28 polypeptide). In some embodiments, the CAR is designated as "X23". In some embodiments, the CAR (e.g., X23) comprises an amino acid sequence that is at least about 85%, about 90%, about 95%, about 96%, about 97%, about 98%, about 99% or about 100% homologous to the amino acid sequence set forth in SEQ ID NO: 164, which is provided in Table 7 below. SEQ ID NO: 164 includes a CD8 leader sequence at amino acids 1 to 18, and is able to bind to CD19 (e.g., human CD19).

**[0364]** In some embodiments, the heterologous cell receptor is a CAR comprising an extracellular antigen-binding domain that binds to a CD19 polypeptide (e.g., a human CD19 polypeptide), a transmembrane domain and a hinge/spacer region derived from a CD28 polypeptide, an intracellular signaling domain comprising a modified CD3ζ polypeptide (e.g., a modified human CD3ζ polypeptide) comprising a native ITAM2, a native BRS1, a native BRS2, a native BRS3, an ITAM1 variant having two loss-of-function mutations, and an ITAM3 variant having two loss-of-function mutations, and a co-stimulatory signaling region comprising a CD28 polypeptide (e.g., a human CD28 polypeptide). In some embodiments, the CAR is designated as "X2X". In some embodiments, the CAR (e.g., X2X) comprises an amino acid sequence that is at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or at least about 100% homologous or 100% identical to the amino acid sequence set forth in SEQ ID NO: 165, which is provided in Table 7 below. SEQ ID NO: 165 includes a CD8 leader sequence at amino acids 1 to 18, and is able to bind to CD19 (e.g., human CD19).

**[0365]** In some embodiments, the heterologous cell receptor is a CAR comprising an extracellular antigen-binding domain that binds to a CD19 polypeptide (e.g., a human CD19 polypeptide), a transmembrane domain and a hinge/spacer region derived from a CD28

polypeptide, an intracellular signaling domain comprising a modified CD3ζ polypeptide (e.g., a modified human CD3ζ polypeptide) comprising a native ITAM1, a native ITAM2, a native BRS1, a native BRS2, a native BRS3, and an ITAM3 variant having two loss-of-function mutations, and a co-stimulatory signaling region comprising a CD28 polypeptide (e.g., a human CD28 polypeptide). In some embodiments, the CAR is designated as "12X". In some embodiments, the CAR (e.g., 12X) comprises an amino acid sequence that is at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or at least about 100% homologous or identical to the amino acid sequence set forth in SEQ ID NO: 166, which is provided in Table 7 below. SEQ ID NO: 166 includes a CD8 leader sequence at amino acids 1 to 18, and is able to bind to CD19 (e.g., human CD19).

**[0366]** In some embodiments, the heterologous cell receptor is a CAR comprising an extracellular antigen-binding domain that binds to a CD19 polypeptide (e.g., human CD19 polypeptide), a transmembrane domain and a hinge/spacer region derived from a CD28 polypeptide, an intracellular signaling domain comprising a modified CD3ζ polypeptide (e.g., a modified human CD3ζ polypeptide) comprising ITAM1, ITAM2, BRS1, BRS2, and a deletion of ITAM3 and a portion of BRS3, and a co-stimulatory signaling region comprising a CD28 polypeptide (e.g., a human CD28 polypeptide), wherein the modified CD3ζ polypeptide comprises a native ITAM1, a native ITAM2, a native BRS1 and a native BRS2, and does not comprise an ITAM3 (native or modified) or a native BRS3. In some embodiments, the CAR is designated as "D3". In some embodiments, the CAR (e.g., D3) comprises an amino acid sequence that is at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or at least about 100% homologous or identical to the amino acid sequence set forth in SEQ ID NO: 167 , which is provided in Table 7 below. SEQ ID NO: 167 includes a CD8 leader sequence at amino acids 1 to 18, and is able to bind to CD19 (e.g., human CD19).

**[0367]** In some embodiments, the heterologous cell receptor is a CAR comprising an extracellular antigen-binding domain that binds to a CD19 polypeptide (e.g., a human CD19 polypeptide), a transmembrane domain and a hinge/spacer region derived from a CD166 polypeptide, an intracellular signaling domain comprising a modified CD3ζ polypeptide (e.g., a modified human CD3ζ polypeptide) comprising a native ITAM1, a native ITAM2, a native

ITAM3, a native BRS1, a native BRS2, and a native BRS3, and a co-stimulatory signaling region comprising a CD28 polypeptide (e.g., a human CD28 polypeptide). In some embodiments, the CAR is designated as "19-166-28z". In some embodiments, the CAR (e.g., 19-166-28z) comprises an amino acid sequence that is at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, at least about 100% homologous or identical to the amino acid sequence set forth in SEQ ID NO: 168, which is provided in Table 7 below. SEQ ID NO: 168 includes a CD8 leader sequence at amino acids 1 to 18, and is able to bind to CD19 (e.g., human CD19).

**[0368]** In some embodiments, the heterologous cell receptor is a CAR comprising an extracellular antigen-binding domain that binds to a CD19 polypeptide (e.g., a human CD19 polypeptide), a transmembrane domain and a hinge/spacer region derived from a CD166 polypeptide, an intracellular signaling domain comprising a modified CD3ζ polypeptide (e.g., a modified human CD3ζ polypeptide) comprising a native ITAM1, a native BRS1, a native BRS2, a native BRS3, an ITAM2 variant having two loss-of-function mutations, and an ITAM3 variant having two loss-of-function mutations, and a co-stimulatory signaling region comprising a CD28 polypeptide (e.g., a human CD28 polypeptide). In some embodiments, the CAR is designated as "19-166-28z 1XX." In some embodiments, the CAR (e.g., 19-166-28z-1XX) comprises an amino acid sequence that is at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, at least about 100% homologous or identical to the amino acid sequence set forth in SEQ ID NO: 169, which is provided in Table 7 below. SEQ ID NO: 169 includes a CD8 leader sequence at amino acids 1 to 18, and is able to bind to CD19 (e.g., human CD19).

**[0369]** In some embodiments, the heterologous cell receptor is a CAR comprising an extracellular antigen-binding domain that binds to a CD19 polypeptide (e.g., human CD19 polypeptide), a transmembrane domain and a hinge/spacer region derived from a CD166 polypeptide, an intracellular signaling domain comprising a modified CD3ζ polypeptide (e.g., a modified human CD3ζ polypeptide) comprising ITAM1, BRS1 and a deletion of ITAM2, ITAM3, BRS2 and BRS3, and a co-stimulatory signaling region comprising a CD28 polypeptide (e.g., a human CD28 polypeptide), wherein the modified CD3ζ polypeptide comprises a native ITAM1 and a native BRS1, and does not comprise an ITAM2 (native or modified), an ITAM3 (native or

modified), a BRS2 (native or modified), or a BRS3 (native or modified). In some embodiments, the CAR is designated as "19-166-28z D23". In some embodiments, the CAR (e.g., 19-166-28z D23) comprises an amino acid sequence that is at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or at least about 100% homologous or identical to the amino acid sequence set forth in SEQ ID NO: 170, which is provided in Table 7 below. SEQ ID NO: 170 includes a CD8 leader sequence at amino acids 1 to 18, and is able to bind to CD19 (e.g., human CD19).

**Table 7 – Exemplary CAR Sequences**

| SEQ ID NO | Description | Sequence |
|---|---|---|
| 159 | Exemplary CAR; 1928z WT construct | MALPVTALLLPLALLLHAEVKLQQSGAELVRPGSSVKIS CKASGYAFSSYWMNWVKQRPGQGLEWIGQIYPGDGDT NYNGKFKGQATLTADKSSSTAYMQLSGLTSEDSAVYFC ARKTISSVVDFYFDYWGQGTTVTVSSGGGGSGGGGSGG GGSDIELTQSPKFMSTSVGDRVSVTCKASQNVGTNVAW YQQKPGQSPKPLIYSATYRNSGVPDRFTGSGSGTDFTLTI TNVQSKDLADYFCQQYNRYPYTSGGGTKLEIKRAAAIE VMYPPPYLDNEKSNGTIIHVKGKHLCPSPLFPGPSKPFW VLVVVGGVLACYSLLVTVAFIIFWVRSKRSRLLHSDYM NMTPRRPGPTRKHYQPYAPPRDFAAYRSRVKFSRSADA PAYQQGQNQLYNELNLGRREEYDVLDKRRGRDPEMGG KPRRKNPQEGLYNELQKDKMAEAYSEIGMKGERRRGK GHDGLYQGLSTATKDTYDALHMQALPPR |
| 160 | Exemplary CAR; 1928z-1XX construct | MALPVTALLLPLALLLHAEVKLQQSGAELVRPGSSVKIS CKASGYAFSSYWMNWVKQRPGQGLEWIGQIYPGDGDT NYNGKFKGQATLTADKSSSTAYMQLSGLTSEDSAVYFC ARKTISSVVDFYFDYWGQGTTVTVSSGGGGSGGGGSGG GGSDIELTQSPKFMSTSVGDRVSVTCKASQNVGTNVAW YQQKPGQSPKPLIYSATYRNSGVPDRFTGSGSGTDFTLTI TNVQSKDLADYFCQQYNRYPYTSGGGTKLEIKRAAAIE VMYPPPYLDNEKSNGTIIHVKGKHLCPSPLFPGPSKPFW VLVVVGGVLACYSLLVTVAFIIFWVRSKRSRLLHSDYM NMTPRRPGPTRKHYQPYAPPRDFAAYRSRVKFSRSADA PAYQQGQNQLYNELNLGRREEYDVLDKRRGRDPEMGG KPRRKNPQEGLFNELQKDKMAEAFSEIGMKGERRRGKG HDGLFQGLSTATKDTFDALHMQALPPR |
| 161 | Exemplary CAR; D12 construct | MALPVTALLLPLALLLHAEVKLQQSGAELVRPGSSVKIS CKASGYAFSSYWMNWVKQRPGQGLEWIGQIYPGDGDT NYNGKFKGQATLTADKSSSTAYMQLSGLTSEDSAVYFC ARKTISSVVDFYFDYWGQGTTVTVSSGGGGSGGGGSGG GGSDIELTQSPKFMSTSVGDRVSVTCKASQNVGTNVAW |

| | | |
|---|---|---|
| | | YQQKPGQSPKPLIYSATYRNSGVPDRFTGSGSGTDFTLTI<br>TNVQSKDLADYFCQQYNRYPYTSGGGTKLEIKRAAAIE<br>VMYPPPYLDNEKSNGTIIHVKGKHLCPSPLFPGPSKPF<br>WVLVVVGGVLACYSLLVTVAFIIFWVRSKRSRLLHSDY<br>MNMTPRRPGPTRKHYQPYAPPRDFAAYRSRVKFSRSAD<br>APAYQQGHDGLYQGLSTATKDTYDALHMQALPPR |
| 162 | Exemplary CAR; D23 construct | MALPVTALLLPLALLLHAEVKLQQSGAELVRPGSSVKIS<br>CKASGYAFSSYWMNWVKQRPGQGLEWIGQIYPGDGDT<br>NYNGKFKGQATLTADKSSSTAYMQLSGLTSEDSAVYFC<br>ARKTISSVVDFYFDYWGQGTTVTVSSGGGGSGGGGSGG<br>GGSDIELTQSPKFMSTSVGDRVSVTCKASQNVGTNVAW<br>YQQKPGQSPKPLIYSATYRNSGVPDRFTGSGSGTDFTLTI<br>TNVQSKDLADYFCQQYNRYPYTSGGGTKLEIKRAAAIE<br>VMYPPPYLDNEKSNGTIIHVKGKHLCPSPLFPGPSKPFW<br>VLVVVGGVLACYSLLVTVAFIIFWVRSKRSRLLHSDYM<br>NMTPRRPGPTRKHYQPYAPPRDFAAYRSRVKFSRSADA<br>PAYQQGQNQLYNELNLGRREEYDVLDKRRGRDP |
| 163 | Exemplary CAR; XX3 construct | MALPVTALLLPLALLLHAEVKLQQSGAELVRPGSSVKIS<br>CKASGYAFSSYWMNWVKQRPGQGLEWIGQIYPGDGDT<br>NYNGKFKGQATLTADKSSSTAYMQLSGLTSEDSAVYFC<br>ARKTISSVVDFYFDYWGQGTTVTVSSGGGGSGGGGSGG<br>GGSDIELTQSPKFMSTSVGDRVSVTCKASQNVGTNVAW<br>YQQKPGQSPKPLIYSATYRNSGVPDRFTGSGSGTDFTLTI<br>TNVQSKDLADYFCQQYNRYPYTSGGGTKLEIKRAAAIE<br>VMYPPPYLDNEKSNGTIIHVKGKHLCPSPLFPGPSKPFW<br>VLVVVGGVLACYSLLVTVAFIIFWVRSKRSRLLHSDYM<br>NMTPRRPGPTRKHYQPYAPPRDFAAYRSRVKFSRSADA<br>PAYQQGQNQLFNELNLGRREEFDVLDKRRGRDPEMGG<br>KPRRKNPQEGLFNELQKDKMAEAFSEIGMKGERRRGKG<br>HDGLYQGLSTATKDTYDALHMQALPPR |
| 164 | Exemplary CAR; X23 construct | MALPVTALLLPLALLLHAEVKLQQSGAELVRPGSSVKIS<br>CKASGYAFSSYWMNWVKQRPGQGLEWIGQIYPGDGDT<br>NYNGKFKGQATLTADKSSSTAYMQLSGLTSEDSAVYFC<br>ARKTISSVVDFYFDYWGQGTTVTVSSGGGGSGGGGSGG<br>GGSDIELTQSPKFMSTSVGDRVSVTCKASQNVGTNVAW<br>YQQKPGQSPKPLIYSATYRNSGVPDRFTGSGSGTDFTLTI<br>TNVQSKDLADYFCQQYNRYPYTSGGGTKLEIKRAAAIE<br>VMYPPPYLDNEKSNGTIIHVKGKHLCPSPLFPGPSKPFW<br>VLVVVGGVLACYSLLVTVAFIIFWVRSKRSRLLHSDYM<br>NMTPRRPGPTRKHYQPYAPPRDFAAYRSRVKFSRSADA<br>PAYQQGQNQLFNELNLGRREEFDVLDKRRGRDPEMGG<br>KPRRKNPQEGLYNELQKDKMAEAYSEIGMKGERRRGK<br>GHDGLYQGLSTATKDTYDALHMQALPPR |

| | | |
|---|---|---|
| 165 | Exemplary CAR; X2X construct | MALPVTALLLPLALLLHAEVKLQQSGAELVRPGSSVKIS<br>CKASGYAFSSYWMNWVKQRPGQGLEWIGQIYPGDGDT<br>NYNGKFKGQATLTADKSSSTAYMQLSGLTSEDSAVYFC<br>ARKTISSVVDFYFDYWGQGTTVTVSSGGGGSGGGGSGG<br>GGSDIELTQSPKFMSTSVGDRVSVTCKASQNVGTNVAW<br>YQQKPGQSPKPLIYSATYRNSGVPDRFTGSGSGTDFTLTI<br>TNVQSKDLADYFCQQYNRYPYTSGGGTKLEIKRAAAIE<br>VMYPPPYLDNEKSNGTIIHVKGKHLCPSPLFPGPSKPFW<br>VLVVVGGVLACYSLLVTVAFIIFWVRSKRSRLLHSDYM<br>NMTPRRPGPTRKHYQPYAPPRDFAAYRSRVKFSRSADA<br>PAYQQGQNQLFNELNLGRREEFDVLDKRRGRDPEMGG<br>KPRRKNPQEGLYNELQKDKMAEAYSEIGMKGERRRGK<br>GHDGLFQGLSTATKDTFDALHMQALPPR |
| 166 | Exemplary CAR; 12X construct | MALPVTALLLPLALLLHAEVKLQQSGAELVRPGSSVKIS<br>CKASGYAFSSYWMNWVKQRPGQGLEWIGQIYPGDGDT<br>NYNGKFKGQATLTADKSSSTAYMQLSGLTSEDSAVYFC<br>ARKTISSVVDFYFDYWGQGTTVTVSSGGGGSGGGGSGG<br>GGSDIELTQSPKFMSTSVGDRVSVTCKASQNVGTNVAW<br>YQQKPGQSPKPLIYSATYRNSGVPDRFTGSGSGTDFTLTI<br>TNVQSKDLADYFCQQYNRYPYTSGGGTKLEIKRAAAIE<br>VMYPPPYLDNEKSNGTIIHVKGKHLCPSPLFPGPSKPFW<br>VLVVVGGVLACYSLLVTVAFIIFWVRSKRSRLLHSDYM<br>NMTPRRPGPTRKHYQPYAPPRDFAAYRSRVKFSRSADA<br>PAYQQGQNQLYNELNLGRREEYDVLDKRRGRDPEMGG<br>KPRRKNPQEGLYNELQKDKMAEAYSEIGMKGERRRGK<br>GHDGLFQGLSTATKDTFDALHMQALPPR |
| 167 | Exemplary CAR; D3 construct | MALPVTALLLPLALLLHAEVKLQQSGAELVRPGSSVKIS<br>CKASGYAFSSYWMNWVKQRPGQGLEWIGQIYPGDGDT<br>NYNGKFKGQATLTADKSSSTAYMQLSGLTSEDSAVYFC<br>ARKTISSVVDFYFDYWGQGTTVTVSSGGGGSGGGGSGG<br>GGSDIELTQSPKFMSTSVGDRVSVTCKASQNVGTNVAW<br>YQQKPGQSPKPLIYSATYRNSGVPDRFTGSGSGTDFTLTI<br>TNVQSKDLADYFCQQYNRYPYTSGGGTKLEIKRAAAIE<br>VMYPPPYLDNEKSNGTIIHVKGKHLCPSPLFPGPSKPFW<br>VLVVVGGVLACYSLLVTVAFIIFWVRSKRSRLLHSDYM<br>NMTPRRPGPTRKHYQPYAPPRDFAAYRSRVKFSRSADA<br>PAYQQGQNQLYNELNLGRREEYDVLDKRRGRDPEMGG<br>KPRRKNPQEGLYNELQKDKMAEAYSEIGMKGERRR |
| 168 | Exemplary CAR; 19-166-28z construct | MALPVTALLLPLALLLHAEVKLQQSGAELVRPGSSVKIS<br>CKASGYAFSSYWMNWVKQRPGQGLEWIGQIYPGDGDT<br>NYNGKFKGQATLTADKSSSTAYMQLSGLTSEDSAVYFC<br>ARKTISSVVDFYFDYWGQGTTVTVSSGGGGSGGGGSGG<br>GGSDIELTQSPKFMSTSVGDRVSVTCKASQNVGTNVAW<br>YQQKPGQSPKPLIYSATYRNSGVPDRFTGSGSGTDFTLTI<br>TNVQSKDLADYFCQQYNRYPYTSGGGTKLEIKRNQLER |

| | | |
|---|---|---|
| | | TVNSLNVSAISIPEHDEADEISDENREKVNDQAKLIVGIV<br>VGLLLAALVAGVVYWLYMKKRSKRSRLLHSDYMNMT<br>PRRPGPTRKHYQPYAPPRDFAAYRKRVKFSRSADAPAY<br>QQGQNQLYNELNLGRREEYDVLDKRRGRDPEMGGKPR<br>RKNPQEGLYNELQKDKMAEAYSEIGMKGERRRGKGHD<br>GLYQGLSTATKDTYDALHMQALPPR |
| 169 | Exemplary CAR; 19-166-28z 1XX construct | MALPVTALLLPLALLLHAEVKLQQSGAELVRPGSSVKIS<br>CKASGYAFSSYWMNWVKQRPGQGLEWIGQIYPGDGDT<br>NYNGKFKGQATLTADKSSSTAYMQLSGLTSEDSAVYFC<br>ARKTISSVVDFYFDYWGQGTTVTVSSGGGGSGGGGSGG<br>GGSDIELTQSPKFMSTSVGDRVSVTCKASQNVGTNVAW<br>YQQKPGQSPKPLIYSATYRNSGVPDRFTGSGSGTDFTLTI<br>TNVQSKDLADYFCQQYNRYPYTSGGGTKLEIKRNQLER<br>TVNSLNVSAISIPEHDEADEISDENREKVNDQAKLIVGIV<br>VGLLLAALVAGVVYWLYMKKRSKRSRLLHSDYMNMT<br>PRRPGPTRKHYQPYAPPRDFAAYRKRVKFSRSADAPAY<br>QQGQNQLYNELNLGRREEYDVLDKRRGRDPEMGGKPR<br>RKNPQEGLFNELQKDKMAEAFSEIGMKGERRRGKGHD<br>GLFQGLSTATKDTFDALHMQALPPR |
| 170 | Exemplary CAR; 19-166-28z D23 construct | MALPVTALLLPLALLLHAEVKLQQSGAELVRPGSSVKIS<br>CKASGYAFSSYWMNWVKQRPGQGLEWIGQIYPGDGDT<br>NYNGKFKGQATLTADKSSSTAYMQLSGLTSEDSAVYFC<br>ARKTISSVVDFYFDYWGQGTTVTVSSGGGGSGGGGSGG<br>GGSDIELTQSPKFMSTSVGDRVSVTCKASQNVGTNVAW<br>YQQKPGQSPKPLIYSATYRNSGVPDRFTGSGSGTDFTLTI<br>TNVQSKDLADYFCQQYNRYPYTSGGGTKLEIKRNQLER<br>TVNSLNVSAISIPEHDEADEISDENREKVNDQAKLIVGIV<br>VGLLLAALVAGVVYWLYMKKRSKRSRLLHSDYMNMT<br>PRRPGPTRKHYQPYAPPRDFAAYRKRVKFSRSADAPAY<br>QQGQNQLYNELNLGRREEYDVLDKRRGRDP |

### 3. Exemplary HLA-independent T Cell Receptors (HITs)

**[0370]** In some embodiments, the heterologous cell receptor is an HLA-independent T cell receptor (HIT). Thus, in some embodiments, the donor template polynucleotide comprises a coding sequence for a HIT. In some embodiments, the HIT exhibits a greater antigen sensitivity than a CAR targeting the same antigen. In some embodiments, the HIT is capable of inducing an immune response when binding to an antigen that has a low density on the surface of a target cell, e.g., a cancer or tumor cell. Non-limiting examples of HITs are discussed in International Patent

Application Publication No. WO 2019/157454 and Mansilla-Soto, Jorge et al., " Nature medicine vol. 28,2 (2022): 345-352. doi:10.1038/s41591-021-01621-1.

**[0371]** In some embodiments, the heterologous cell receptor is a HIT comprising an extracellular antigen-binding domain derived from an antibody fragment such as an scFv or an Fab. In some embodiments, the extracellular antigen-binding domain also comprises a constant domain. In some embodiments, the extracellular antigen-binding domain is capable of dimerizing with another extracellular antigen-binding domain (e.g., forming a fragment variable (Fv)), wherein the dimerized antigen-binding domains (e.g., an Fv) specifically bind to an antigen, e.g., a tumor antigen or a pathogen antigen.

**[0372]** In some embodiments, the HIT extracellular antigen-binding domain comprises a heavy chain variable region (VH) and/or a light chain variable region (VL) of an antibody, wherein the VH or the VL is capable of dimerizing with another extracellular antigen binding domain comprising a VL or a VH (e.g., forming a fragment variable (Fv)).

**[0373]** In some embodiments, a TCR's HLA-restricted Vα-Vβ pair is replaced by heterologous VL-VH binding domains for heterologous antigen-binding specificity. In some embodiments, the VL-VH binding domains confer specificity for target polypeptide. The engineered VH-Cβ and VL-Cα chains can associate to form the HIT heterodimer with the VH and VL portions forming the fragment variable Fv.

**[0374]** In some embodiments, a TCR or CAR anti-CD19 heavy chain sequence is covalently linked to a TCR Cβ sequence (VHCβ). In some embodiments, a TCR or CAR anti-CD19 light chain sequence to the TCR Cα sequence (VLCα).

**[0375]** In some embodiments, the HIT extracellular antigen-binding domain comprises an anti-CD19 VH amino acid sequence that is at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% identical to the VH amino acid sequence of the CD19 monoclonal antibody SJ25C1 disclosed in International Patent Application No. WO 2019/157454. (see SEQ ID NO: 7 and 44). In some embodiments, the HIT extracellular antigen-binding domain comprises an anti-CD19 VL amino acid sequence that is at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100%

identical to the VL amino acid sequence of the CD19 monoclonal antibody SJ25C1 disclosed in International Patent Application No. WO 2019/157454 (see SEQ ID NO: 8 and 45). In some embodiments, the HIT comprises an scFv sequence that is at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% identical to the scFv amino acid sequence of the CD19 monoclonal antibody SJ25C1 disclosed in International Patent Application No. WO 2019/157454 (see SEQ ID NO: 10).

**[0376]** In some embodiments, the HIT comprises a VH CDR1 sequence that is at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% identical to the amino acid sequence GYAFSS (SEQ ID NO: 171).

**[0377]** In some embodiments, the HIT comprises a VH CDR2 sequence that is at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% identical to the amino acid sequence YPGDGD (SEQ ID NO: 172).

**[0378]** In some embodiments, the HIT comprises a VH CDR3 sequence that is at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% identical to the amino acid sequence KTISSVVDF (SEQ ID NO: 173).

**[0379]** In some embodiments, the HIT comprises a VL CDR1 sequence that is at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% identical to the amino acid sequence KASQNVGTNVA (SEQ ID NO: 174).

**[0380]** In some embodiments, the HIT comprises a VL CDR2 sequence that is at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% identical to the amino acid sequence SATYRN (SEQ ID NO: 175).

**[0381]** In some embodiments, the HIT comprises a VL CDR3 sequence that is at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about

97%, at least about 98%, at least about 99%, or 100% identical to the amino acid sequence QQYNTYPYT (SEQ ID NO: 176).

**[0382]** In some embodiments, the HIT comprises a VH-Cβ sequence that that is at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% identical to the amino acid sequence as set forth in SEQ ID NO: 177.

EVKLQESGPGLVAPSQSLSVTCTVSGVSLPDYGVSWIRQPPRKGLEWLGVIWGSETTYY NSALKSRLTIIKDNSKSQVFLKMNSLQTDDTAIYYCAKHYYYGGSYAMDYWGQGTSVT VLEDLKNVFPPEVAVFEPSEAEISHTQKATLVCLATGFYPDHVELSWWVNGKEVHSGVS TDPQPLKEQPALNDSRYCLSSRLRVSATFWQNPRNHFRCQVQFYGLSENDEWTQDRAK PVTQIVSAEAWGRADCGFTSESYQQGVLSATILYEILLGKATLYAVLVSALVLMAMVKR KDSRG (SEQ ID NO: 177)

**[0383]** In some embodiments, the HIT comprises a VL-Cα sequence that that is at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% identical to the amino acid sequence as set forth in SEQ ID NO: 178.

DIQMTQTTSSLSASLGDRVTISCRASQDISKYLNWYQQKPDGTVKLLIYHTSRLHSGVPS RFSGSGSGTDYSLTISNLEQEDIATYFCQQGNTLPYTFGGGTKLEIIPNIQNPDPAVYQLR DSKSSDKSVCLFTDFDSQTNVSQSKDSDVYITDKTVLDMRSMDFKSNSAVAWSNKSDF ACANAFNNSIIPEDTFFPSP (SEQ ID NO: 178)

**[0384]** In some embodiments, a TCR or CAR anti-CD22 heavy chain sequence is covalently linked to a TCR Cβ sequence (VHCβ). In some embodiments, a TCR or CAR anti-CD22 light chain sequence to the TCR Cα sequence (VLCα). In some embodiments, the HIT extracellular antigen-binding domain comprises anti-CD22 $V_L$-$V_H$ sequences derived from CARs are previously described in Priceman SJ et al., Oncoimmunology 7, e1380764, doi:10.1080/2162402X.2017.1380764 (2018).

**[0385]** In some embodiments, a TCR or CAR anti-BCMA heavy chain sequence is covalently linked to a TCR Cβ sequence (VHCβ). In some embodiments, a TCR or CAR anti-BCMA light

chain sequence to the TCR Cα sequence (VLCα). In some embodiments, the HIT extracellular antigen-binding domain comprises anti-BCMA $V_L$-$V_H$ sequences derived from CARs are previously described in Priceman SJ et al., Oncoimmunology 7, e1380764, doi:10.1080/2162402X.2017.1380764 (2018).

**[0386]** In some embodiments, the extracellular antigen-binding domain further comprises a constant domain. In some embodiments, the constant domain comprises a hinge/spacer region and a transmembrane domain. In some embodiments, the constant domain is capable of forming a homodimer or a heterodimer with another constant domain. In some embodiments, the constant domain dimerizes through one or more disulfide-links. In some embodiments, the antigen binding chain is capable of forming a trimer or oligomer with one or more identical or different constant domains. In some non-limiting embodiments, the constant domain comprises a T cell receptor constant region, e.g., T cell receptor alpha constant region (TRAC), T cell receptor beta constant region (TRBC, e.g., TRBC1 or TRBC2), T cell receptor gamma constant region (TRGC, e.g., TRGC1 or TRGC2), T cell receptor delta constant region (TRDC) or any variants or functional fragments thereof.

**[0387]** In some embodiments, the HIT constant domain comprises a native or modified TRAC polypeptide. In certain embodiments, the TRAC polypeptide comprises an amino acid sequence that is at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99% or 100% identical to the amino acid sequence set forth in SEQ ID NO: 179, which is provided below), or fragments thereof, and/or may optionally comprise up to one or up to two or up to three conservative amino acid substitutions.

IQNPDPAVYQLRDSKSSDKSVCLFTDFDSQTNVSQSKDSDVYITDKTVLDMRSMDFKSN SAVAWSNKSDFACANAFNNSIIPEDTFFPSPESSCDVKLVEKSFETDTNLNFQNLSVIGFR ILLLKVAGFNLLMTLRLWSS (SEQ ID NO: 179)

**[0388]** In certain embodiments, the TRAC polypeptide has an amino acid sequence that is at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99% or 100% identical to the amino acid sequence encoded by a transcript expressed by the gene of NCBI Genbank ID: 28755, NG_001332.3, range 925603 to 930229 (SEQ ID NO: 180, which is provided below), or fragments

thereof, and/or may optionally comprise up to one or up to two or up to three conservative amino acid substitutions.

ATATCCAGAACCCTGACCCTGCCGTGTACCAGCTGAGAGACTCTAAATCCAGTGACA
AGTCTGTCTGCCTATTCACCGATTTTGATTCTCAAACAAATGTGTCACAAAGTAAGG
ATTCTGATGTGTATATCACAGACAAAACTGTGCTAGACATGAGGTCTATGGACTTCA
AGAGCAACAGTGCTGTGGCCTGGAGCAACAAATCTGACTTTGCATGTGCAAACGCC
TTCAACAACAGCATTATTCCAGAAGACACCTTCTTCCCCAGCCCAGGTAAGGGCAGC
TTTGGTGCCTTCGCAGGCTGTTTCCTTGCTTCAGGAATGGCCAGGTTCTGCCCAGAGC
TCTGGTCAATGATGTCTAAAACTCCTCTGATTGGTGGTCTCGGCCTTATCCATTGCCA
CCAAAACCCTCTTTTTACTAAGAAACAGTGAGCCTTGTTCTGGCAGTCCAGAGAATG
ACACGGGAAAAAAGCAGATGAAGAGAAGGTGGCAGGAGAGGGCACGTGGCCCAGC
CTCAGTCTCTCCAACTGAGTTCCTGCCTGCCTGCCTTTGCTCAGACTGTTTGCCCCTT
ACTGCTCTTCTAGGCCTCATTCTAAGCCCCTTCTCCAAGTTGCCTCTCCTTATTTCTCC
CTGTCTGCCAAAAAATCTTTCCCAGCTCACTAAGTCAGTCTCACGCAGTCACTCATT
AACCCACCAATCACTGATTGTGCCGGCACATGAATGCACCAGGTGTTGAAGTGGAG
GAATTAAAAAGTCAGATGAGGGGTGTGCCCAGAGGAAGCACCATTCTAGTTGGGGG
AGCCCATCTGTCAGCTGGGAAAAGTCCAAATAACTTCAGATTGGAATGTGTTTTAAC
TCAGGGTTGAGAAAACAGCTACCTTCAGGACAAAAGTCAGGGAAGGGCTCTCTGAA
GAAATGCTACTTGAAGATACCAGCCCTACCAAGGGCAGGGAGAGGACCCTATAGAG
GCCTGGGACAGGAGCTCAATGAGAAAGGAGAAGAGCAGCAGGCATGAGTTGAATG
AAGGAGGCAGGGCCGGGTCACAGGGCCTTCTAGGCCATGAGAGGGTAGACAGTATT
CTAAGGACGCCAGAAAGCTGTTGATCGGCTTCAAGCAGGGGAGGGACACCTAATTT
GCTTTTCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGAGATGGAGTTTTGCTCTTGTTGC
CCAGGCTGGAGTGCAATGGTGCATCTTGGCTCACTGCAACCTCCGCCTCCCAGGTTC
AAGTGATTCTCCTGCCTCAGCCTCCCGAGTAGCTGAGATTACAGGCACCCGCCACCA
TGCCTGGCTAATTTTTTGTATTTTTAGTAGAGACAGGGTTTCACTATGTTGGCCAGGC
TGGTCTCGAACTCCTGACCTCAGGTGATCCACCCGCTTCAGCCTCCCAAAGTGCTGG
GATTACAGGCGTGAGCCACCACACCCGGCCTGCTTTTCTTAAAGATCAATCTGAGTG
CTGTACGGAGAGTGGGTTGTAAGCCAAGAGTAGAAGCAGAAAGGGAGCAGTTGCA
GCAGAGAGATGATGGAGGCCTGGGCAGGGTGGTGGCAGGGAGGTAACCAACACCA
TTCAGGTTTCAAAGGTAGAACCATGCAGGGATGAGAAAGCAAAGAGGGGATCAAG
GAAGGCAGCTGGATTTTGGCCTGAGCAGCTGAGTCAATGATAGTGCCGTTTACTAAG
AAGAAACCAAGGAAAAAATTTGGGGTGCAGGGATCAAAACTTTTTGGAACATATGA
AAGTACGTGTTTATACTCTTTATGGCCCTTGTCACTATGTATGCCTCGCTGCCTCCAT
TGGACTCTAGAATGAAGCCAGGCAAGAGCAGGGTCTATGTGTGATGGCACATGTGG
CCAGGGTCATGCAACATGTACTTTGTACAAACAGTGTATATTGAGTAAATAGAAATG
GTGTCCAGGAGCCGAGGTATCGGTCCTGCCAGGGCCAGGGGCTCTCCCTAGCAGGT
GCTCATATGCTGTAAGTTCCCTCCAGATCTCTCCACAAGGAGGCATGGAAAGGCTGT
AGTTGTTCACCTGCCCAAGAACTAGGAGGTCTGGGGTGGGAGAGTCAGCCTGCTCTG
GATGCTGAAAGAATGTCTGTTTTTCCTTTTAGAAAGTTCCTGTGATGTCAAGCTGGTC
GAGAAAAGCTTTGAAACAGGTAAGACAGGGGTCTAGCCTGGGTTTGCACAGGATTG
CGGAAGTGATGAACCCGCAATAACCCTGCCTGGATGAGGGAGTGGGAAGAAATTAG
TAGATGTGGGAATGAATGATGAGGAATGGAAACAGCGGTTCAAGACCTGCCCAGAG
CTGGGTGGGGTCTCTCCTGAATCCCTCTCACCATCTCTGACTTTCCATTCTAAGCACT
TTGAGGATGAGTTTCTAGCTTCAATAGACCAAGGACTCTCTCCTAGGCCTCTGTATTC

CTTTCAACAGCTCCACTGTCAAGAGAGCCAGAGAGAGCTTCTGGGTGGCCCAGCTGT
GAAATTTCTGAGTCCCTTAGGGATAGCCCTAAACGAACCAGATCATCCTGAGGACA
GCCAAGAGGTTTTGCCTTCTTTCAAGACAAGCAACAGTACTCACATAGGCTGTGGGC
AATGGTCCTGTCTCTCAAGAATCCCCTGCCACTCCTCACACCCACCCTGGGCCCATA
TTCATTTCCATTTGAGTTGTTCTTATTGAGTCATCCTTCCTGTGGTAGCGGAACTCAC
TAAGGGGCCCATCTGGACCCGAGGTATTGTGATGATAAATTCTGAGCACCTACCCCA
TCCCCAGAAGGGCTCAGAAATAAAATAAGAGCCAAGTCTAGTCGGTGTTTCCTGTCT
TGAAACACAATACTGTTGGCCCTGGAAGAATGCACAGAATCTGTTTGTAAGGGGAT
ATGCACAGAAGCTGCAAGGGACAGGAGGTGCAGGAGCTGCAGGCCTCCCCCACCCA
GCCTGCTCTGCCTTGGGGAAAACCGTGGGTGTGTCCTGCAGGCCATGCAGGCCTGGG
ACATGCAAGCCCATAACCGCTGTGGCCTCTTGGTTTTACAGATACGAACCTAAACTT
TCAAAACCTGTCAGTGATTGGGTTCCGAATCCTCCTCCTGAAAGTGGCCGGGTTTAA
TCTGCTCATGACGCTGCGGCTGTGGTCCAGCTGAGGTGAGGGGCCTTGAAGCTGGGA
GTGGGGTTTAGGGACGCGGGTCTCTGGGTGCATCCTAAGCTCTGAGAGCAAACCTCC
CTGCAGGGTCTTGTTTTAAGTCCAAAGCCTGAGCCCACCAAACTCTCCTACTTCTTCC
TGTTACAAATTCCTCTTGTGCAATAATAATGGCCTGAAACGCTGTAAAATATCCTCA
TTTCAGCCGCCTCAGTTGCACTTCTCCCCTATGAGGTAGGAAGAACAGTTGTTTAGA
AACGAAGAAACTGAGGCCCCACAGCTAATGAGTGGAGGAAGAGAGACACTTGTGTA
CACCACATGCCTTGTGTTGTACTTCTCTCACCGTGTAACCTCCTCATGTCCTCTCTCC
CCAGTACGGCTCTCTTAGCTCAGTAGAAAGAAGACATTACACTCATATTACACCCCA
ATCCTGGCTAGAGTCTCCGCACCCTCCTCCCCCAGGGTCCCCAGTCGTCTTGCTGAC
AACTGCATCCTGTTCCATCACCATCAAAAAAAAACTCCAGGCTGGGTGCGGGGGCT
CACACCTGTAATCCCAGCACTTTGGGAGGCAGAGGCAGGAGGAGCACAGGAGCTGG
AGACCAGCCTGGGCAACACAGGGAGACCCCGCCTCTACAAAAAGTGAAAAAATTAA
CCAGGTGTGGTGCTGCACACCTGTAGTCCCAGCTACTTAAGAGGCTGAGATGGGAG
GATCGCTTGAGCCCTGGAATGTTGAGGCTACAATGAGCTGTGATTGCGTCACTGCAC
TCCAGCCTGGAAGACAAAGCAAGATCCTGTCTCAAATAATAAAAAAAATAAGAACT
CCAGGGTACATTTGCTCCTAGAACTCTACCACATAGCCCCAAACAGAGCCATCACCA
TCACATCCCTAACAGTCCTGGGTCTTCCTCAGTGTCCAGCCTGACTTCTGTTCTTCCT
CATTCCAGATCTGCAAGATTGTAAGACAGCCTGTGCTCCCTCGCTCCTTCCTCTGCAT
TGCCCCTCTTCTCCCTCTCCAAACAGAGGGAACTCTCCTACCCCCAAGGAGGTGAAA
GCTGCTACCACCTCTGTGCCCCCCCGGCAATGCCACCAACTGGATCCTACCCGAATT
TATGATTAAGATTGCTGAAGAGCTGCCAAACACTGCTGCCACCCCCTCTGTTCCCTT
ATTGCTGCTTGTCACTGCCTGACATTCACGGCAGAGGCAAGGCTGCTGCAGCCTCCC
CTGGCTGTGCACATTCCCTCCTGCTCCCCAGAGACTGCCTCCGCCATCCCACAGATG
ATGGATCTTCAGTGGGTTCTCTTGGGCTCTAGGTCCTGCAGAATGTTGTGAGGGGTT
TATTTTTTTTTAATAGTGTTCATAAAGAAATACATAGTATTCTTCTTCTCAAGACGTG
GGGGGAAATTATCTCATTATCGAGGCCCTGCTATGCTGTGTATCTGGGCGTGTTGTA
TGTCCTGCTGCCGATGCCTTC (SEQ ID NO: 180)

**[0389]** In some embodiments, the HIT constant domain comprises a native or modified TRBC peptide. In some embodiments, the HIT constant domain comprises a native or modified TRBC2 peptide. In some embodiments, the TRBC2 polypeptide comprises an amino acid sequence that is at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99% or 100% identical to the amino acid

sequence set forth in SEQ ID NO: 181, which is provided below, or fragments thereof, and/or may optionally comprise up to one or up to two or up to three conservative amino acid substitutions.

DLKNVFPPEVAVFEPSEAEISHTQKATLVCLATGFYPDHVELSWWVNGKEVHSGVSTDP
QPLKEQPALNDSRYCLSSRLRVSATFWQNPRNHFRCQVQFYGLSENDEWTQDRAKPVT
QIVSAEAWGRADCGFTSESYQQGVLSATILYEILLGKATLYAVLVSALVLMAMVKRKD
SRG (SEQ ID NO: 181)

**[0390]** In certain embodiments, the HIT constant domain comprises a native or modified TRBC1 peptide. In certain embodiments, the TRBC1 polypeptide comprises an amino acid sequence that is at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99% or 100% identical to the amino acid sequence set forth in SEQ ID NO: 182, which is provided below, or fragments thereof, and/or may optionally comprise up to one or up to two or up to three conservative amino acid substitutions.

DLNKVFPPEVAVFEPSEAEISHTQKATLVCLATGFFPDHVELSWWWGKEVHSGVSTDPQ
PLKEQPALNDSRYCLSSRLRVSATFWQNPRNHFRCQVQFYGLSENDEWTQDRAKPVTQ
IVSAEAWGRADCGFTSVSYQQGVLSATILYEILLGKATLYAVLVSALVLMAMVKRKDF
(SEQ ID NO: 182)

**[0391]** In certain embodiments, the TRBC polypeptide has an amino acid sequence that is at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99% or 100% identical to the amino acid sequence encoded by a transcript expressed by a gene of NCBI Genbank ID: 28639, NG_001333.2, range 645749 to 647196 (TRBC1, SEQ ID NO: 183), NCBI Genbank ID: 28638, NG 001333.2 range 655095 to 656583 (TRBC2, SEQ ID NO: 184) or fragments thereof, and/or may optionally comprise up to one or up to two or up to three conservative amino acid substitutions.

AGGACCTGAACAAGGTGTTCCCACCCGAGGTCGCTGTGTTTGAGCCATCAGAAGCA
GAGATCTCCCACACCCAAAAGGCCACACTGGTGTGCCTGGCCACAGGCTTCTTCCCC
GACCACGTGGAGCTGAGCTGGTGGGTGAATGGGAAGGAGGTGCACAGTGGGGTCAG
CACAGACCCGCAGCCCCTCAAGGAGCAGCCCGCCCTCAATGACTCCAGATACTGCC
TGAGCAGCCGCCTGAGGGTCTCGGCCACCTTCTGGCAGAACCCCCGCAACCACTTCC
GCTGTCAAGTCCAGTTCTACGGGCTCTCGGAGAATGACGAGTGGACCCAGGATAGG
GCCAAACCCGTCACCCAGATCGTCAGCGCCGAGGCCTGGGGTAGAGCAGGTGAGTG
GGGCCTGGGGAGATGCCTGGAGGAGATTAGGTGAGACCAGCTACCAGGGAAAATG
GAAAGATCCAGGTAGCAGACAAGACTAGATCCAAAAAGAAAGGAACCAGCGCACA
CCATGAAGGAGAATTGGGCACCTGTGGTTCATTCTTCTCCCAGATTCTCAGCCCAAC
AGAGCCAAGCAGCTGGGTCCCCTTTCTATGTGGCCTGTGTAACTCTCATCTGGGTGG
TGCCCCCCATCCCCCTCAGTGCTGCCACATGCCATGGATTGCAAGGACAATGTGGCT

GACATCTGCATGGCAGAAGAAAGGAGGTGCTGGGCTGTCAGAGGAAGCTGGTCTGG
GCCTGGGAGTCTGTGCCAACTGCAAATCTGACTTTACTTTTAATTGCCTATGAAAAT
AAGGTCTCTCATTTATTTTCCTCTCCCTGCTTTCTTTCAGACTGTGGCTTTACCTCGGG
TAAGTAAGCCCTTCCTTTTCCTCTCCCTCTCTCATGGTTCTTGACCTAGAACCAAGGC
ATGAAGAACTCACAGACACTGGAGGGTGGAGGGTGGGAGAGACCAGAGCTACCTGT
GCACAGGTACCCACCTGTCCTTCCTCCGTGCCAACAGTGTCCTACCAGCAAGGGGTC
CTGTCTGCCACCATCCTCTATGAGATCCTGCTAGGGAAGGCCACCCTGTATGCTGTG
CTGGTCAGCGCCCTTGTGTTGATGGCCATGGTAAGCAGGAGGGCAGGATGGGGCCA
GCAGGCTGGAGGTGACACACTGACACCAAGCACCCAGAAGTATAGAGTCCCTGCCA
GGATTGGAGCTGGGCAGTAGGGAGGGAAGAGATTTCATTCAGGTGCCTCAGAAGAT
AACTTGCACCTCTGTAGGATCACAGTGGAAGGGTCATGCTGGGAAGGAGAAGCTGG
AGTCACCAGAAAACCCAATGGATGTTGTGATGAGCCTTACTATTTGTGTGGTCAATG
GGCCCTACTACTTTCTCTCAATCCTCACAACTCCTGGCTCTTAATAACCCCCAAAACT
TTCTCTTCTGCAGGTCAAGAGAAAGGATTTCTGA (SEQ ID NO: 183)

AGGACCTGAAAAACGTGTTCCCACCCGAGGTCGCTGTGTTTGAGCCATCAGAAGCA
GAGATCTCCCACACCCAAAAGGCCACACTGGTATGCCTGGCCACAGGCTTCTACCCC
GACCACGTGGAGCTGAGCTGGTGGGTGAATGGGAAGGAGGTGCACAGTGGGGTCAG
CACAGACCCGCAGCCCCTCAAGGAGCAGCCCGCCCTCAATGACTCCAGATACTGCC
TGAGCAGCCGCCTGAGGGTCTCGGCCACCTTCTGGCAGAACCCCCGCAACCACTTCC
GCTGTCAAGTCCAGTTCTACGGGCTCTCGGAGAATGACGAGTGGACCCAGGATAGG
GCCAAACCCGTCACCCAGATCGTCAGCGCCGAGGCCTGGGGTAGAGCAGGTGAGTG
GGGCCTGGGGAGATGCCTGGAGGAGATTAGGTGAGACCAGCTACCAGGGAAAATG
GAAAGATCCAGGTAGCGGACAAGACTAGATCCAGAAGAAAGCCAGAGTGGACAAG
GTGGGATGATCAAGGTTCACAGGGTCAGCAAAGCACGGTGTGCACTTCCCCCACCA
AGAAGCATAGAGGCTGAATGGAGCACCTCAAGCTCATTCTTCCTTCAGATCCTGACA
CCTTAGAGCTAAGCTTTCAAGTCTCCCTGAGGACCAGCCATACAGCTCAGCATCTGA
GTGGTGTGCATCCCATTCTCTTCTGGGGTCCTGGTTTCCTAAGATCATAGTGACCACT
TCGCTGGCACTGGAGCAGCATGAGGGAGACAGAACCAGGGCTATCAAAGGAGGCTG
ACTTTGTACTATCTGATATGCATGTGTTTGTGGCCTGTGAGTCTGTGATGTAAGGCTC
AATGTCCTTACAAAGCAGCATTCTCTCATCCATTTTTCTTCCCCTGTTTTCTTTCAGAC
TGTGGCTTCACCTCCGGTAAGTGAGTCTCTCCTTTTTCTCTCTATCTTTCGCCGTCTCT
GCTCTCGAACCAGGGCATGGAGAATCCACGGACACAGGGGCGTGAGGGAGGCCAG
AGCCACCTGTGCACAGGTGCCTACATGCTCTGTTCTTGTCAACAGAGTCTTACCAGC
AAGGGGTCCTGTCTGCCACCATCCTCTATGAGATCTTGCTAGGGAAGGCCACCTTGT
ATGCCGTGCTGGTCAGTGCCCTCGTGCTGATGGCCATGGTAAGGAGGAGGGTGGGA
TAGGGCAGATGATGGGGGCAGGGGATGGAACATCACACATGGGCATAAAGGAATCT
CAGAGCCAGAGCACAGCCTAATATATCCTATCACCTCAATGAAACCATAATGAAGC
CAGACTGGGGAGAAAATGCAGGGAATATCACAGAATGCATCATGGGAGGATGGAG
ACAACCAGCGAGCCCTACTCAAATTAGGCCTCAGAGCCCGCCTCCCCTGCCCTACTC
CTGCTGTGCCATAGCCCCTGAAACCCTGAAAATGTTCTCTCTTCCACAGGTCAAGAG
AAAGGATTCCAGAGGCTAG (SEQ ID NO: 184)

**[0392]** In some embodiments, the HIT constant domain comprises a native or modified TRGC peptide. In some embodiments, the HIT constant domain comprises a native or modified TRGC1 peptide. In some embodiments, the TRGC1 polypeptide has an amino acid sequence that is at least

about 85%, about 90%, about 95%, about 96%, about 97%, about 98%, about 99% or 100% homologous or identical to the amino acid sequence set forth in SEQ ID NO: 185, which is provided below.

DKQLDADVSPKPTIFLPSIAETKLQKAGTYLCLLEKFFPDVIKIHWQEKKSNTILGSQEGN
TMKTNDTYMKFSWLTVPEKSLDKEHRCIVRHENNKNGVDQEIIFPPIKTDVITMDPKDN
CSKDANDTLLLQLTNTSAYYMYLLLLLKSVVYFAIITCCLLRRTAFCCNGEKS
(SEQ ID NO: 185)

**[0393]** In some embodiments, the HIT constant domain comprises a native or modified TRGC2 peptide. In some embodiments, the TRGC2 polypeptide has an amino acid sequence that is at least about 85%, about 90%, about 95%, about 96%, about 97%, about 98%, about 99% or 100% homologous or identical to the amino acid sequence set forth in SEQ ID NO: 186, which is provided below.

DKQLDADVSPKPTIFLPSIAETKLQKAGTYLCLLEKFFPDIIKIHWQEKKSNTILGSQEGNT
MKTNDTYMKFSWLTVPEESLDKEHRCIVRHENNKNGIDQEIIFPPIKTDVTTVDPKYNYS
KDANDVITMDPKDNWSKDANDTLLLQLTNTSAYYTYLLLLLKSWYFAIITCCLLRRTAF
CCNGEKS (SEQ ID NO: 186)

**[0394]** In some embodiments, the TRGC polypeptide has an amino acid sequence that is at least about 85%, about 90%, about 95%, about 96%, about 97%, about 98%, about 99% or 100% homologous or identical to the amino acid sequence encoded by a transcript expressed by a gene of NCBI Genbank ID: 6966, NG_001336.2, range 108270 to 113860 (TRGC1, SEQ ID NO: 187), NCBI Genbank ID: 6967, NG_001336.2, range 124376 to 133924 (TRGC2, SEQ ID NO: 188) or fragments thereof, and/or may optionally comprise up to one or up to two or up to three conservative amino acid substitutions.

ATAAACAACTTGATGCAGATGTTTCCCCCAAGCCCACTATTTTTCTTCCTTCAATTGC
TGAAACAAAGCTCCAGAAGGCTGGAACATACCTTTGTCTTCTTGAGAAATTTTTCCC
TGATGTTATTAAGATACATTGGCAAGAAAAGAAGAGCAACACGATTCTGGGATCCC
AGGAGGGGAACACCATGAAGACTAACGACACATACATGAAATTTAGCTGGTTAACG
GTGCCAGAAAAGTCACTGGACAAAGAACACAGATGTATCGTCAGACATGAGAATAA
TAAAAACGGAGTTGATCAAGAAATTATCTTTCCTCCAATAAAGACAGGTATGTGTTT
ACGCATATCATCTGTCAGAACACTTCTTTGAAAGTGAATGCTGCATTTTTTCCTTTCA
GTATTAATGAAAAACAAACATAAATCTTTCTTAAATATTGTTACATTTAATGGTAGC
ATAAATGCCCTGCTACTTTTCTATAGAATTAAAATGGTATAGGTTTTGGAGAAAACA
AAATTGAAAAAGTTACTGAAGGTTTGTCAGCCTCAGCTCCATTATCCAAAATAAGAA
AGTCACGTGCTGGTTTTTAGGGTTGTTAGATGGATTAAAGAAACAACATACACAGAA
GCATCTAGCAACGTGACACGTGGTAAACGCTCAAAAAGTGTTCTCCCTTCTTTTGAT
GACTTTACTTGATCAGGAAATAACATATATATGTCTTTCAGGAATGTTCTGCCCAAG

CAGGAGAGTCACTCACCTCAATCTTGCTACCCACAAAGTTTAACCTAAAAACAACG
GGTTCATTGTTGACAAAATGATGTTTATCTGTTGTTGACAGAATGATGTTTATCTAAA
AACAGTTCCAATTTTCTATTTCCTTTGCTGAGACACAAAGGGGAGGCAAATGTGCAA
AGCTTGAGGGTAGTCTTACCACTGTGCTTAAGTGTTCTGATTTTTCTAGTGATCAGGG
CAAAATAAAAAGTATAGTAAGTTCCAAGGCAGTGAATATTATACAGGAGAGAAGTT
ACAGTTTTATAATGTGTTTTCCTTTACACTAAATTCTAAAAGTAAAAAGTCTTTTTTT
TTTTTTGACAGAGTTTCACTCTTGTTGCCCAAGCAGGTGTGCTATGGTATGATCTCAG
CTCACTGCAACCTCCACCTCCCGGGTTCAAGTGATTCTCTTACTTCAGCCTCCCGACA
GGCTGGGATTGCAGGCGCCTGCCACCACACCTGGCTAATTTTTGTGTTTTTAGTAGA
GATGGGGTTTCACCATGTTGGCCAGGCTGGTCTCAAATTCCTGACCTCAAGTGATCC
ATCCACCTCGGCCTCCAAGTGCTGGGATTATGGGCGTCAGCCACTGTGCCCAGCCTA
AAAGTAAAATGTCTTTCATGAGCTTCCCAAGGCAGCTACGTTAAGGAGGACACTTCT
CTTAATGTCATTCTACAGTAGATTTCTAATGCTCTTTCTTGGAAGTTTGTTTTTCTGAG
AAAAGCTAAAAATATAACATGGAAGTGATCATATTATATAATCAATGAAGTGCTTTT
CAAGGAGATAAAACTAATCTGGTCCACACTTGCAACCAACCTTGATTGAGAGAGAG
AGAGAACTCAGGATACACTTGAAGATTTTATTATGGGGAACAGTTACTTTATTCTTT
TTACCTCAATCAATGCATGGAAATAAGTGATAGTCATTTTCATTTATCTTTTAATAAA
TGAAGTCACCATGAGGAAAATAAAAAGACATTGAAAACCCATTAAAGTCAGCCCTT
AAAGATATTTGGACATGCAGACTTGATAACTAACGTTTGCATTCTTGAGACTTACCC
AAAACCCATACCTCAAGTCCAAGTTTTTAGAATTCATGAAATAAAGATCTCAGTGAG
TGCATAAAATTGCGCACCAGAATCATATCCGTATAGACAAGAACACATCTACTAGA
AAAATAATAAACCAACACACCAATGCAACTGTGTTTTCTTCTGTTTTAAAGTATGTT
GTCTTTGTATGCATGTTTGCTTCTTCCTTTTTTTTTTTAACATCACAGATAAATTCAAC
TCTCACCTCAGGTTTTATTGAGAGAACTGTCAATGTGACTTGGCCTCTGTCTTTCTAG
TCCCAGAAAGAATTGCACTGAAATCTGAGCTCCTGTAATAAAAACAACCATTTGCTG
AGAGTAATTAACATACTGAAAGAGATTTTCTTAGAGTACACAATGGTGACATTATAT
TGCCTCTTTATAAATAACTTTCTATCTATTTCTGTGGATTATTCCTACAAAGTACTTTT
CATATGTCCAATTTCTTTTCTTCCCCTACAACTACTGTCTGAATACTGGCTCTGCTATT
TGCTGATATGATTCTCGGCAAGTTGCCTGCACTTTTTAAACTTTATTTCCTCATTCAG
AACATGGGGCCATACATAATACAACTCACTTCAGTGTTATTGGGGAATTAAACAAA
AAATGCATGGGAAGCATTTAACATAGTGCCTGACACAATAATGAGTACTCAGTAGA
TGTTAGCTTTTATTAATATTGTTGTTGTTATGTCCAGAAACACTATACCTCCAGAAAA
TCATGGGTACTTGCTGGGGACATTGGGGATATGCATGATTTGGAAAAGAATGACTGC
TTTTTTTGCTTAGATGAGAAATTTTTCTAAGCCAGACTCCTTCAAATATGTAAGATTC
TGTTGTGGATTCAAGGACTGAAAGAATTCTTGGCCGAGTGTGGTGGCTTATCCCTGT
AATCCCAGCATTTTGTGAGGACAAGGCAGGAAGATTGCTTGAGTCCAGGAGTTTGA
AACCAGCCTGCGCAACATGGCGAAACCCTGTCTCTACAAAAAATACAAACATTAGC
TCGGAGTGAGTGCTGACATGTGCCTGTACTCCCAGCTACTCAGAAGGCTGAGATGGG
AGGATCTCATGAGCCTGGGGAGTTTGAGGCTTCAGTGAGCCGTGATGACACCGTACT
ATACTCCACTCCAGCCTGGGTGACAGTGAGACCCTGCCTCAAAAAACAAACAAACA
AACAAACAAAACAAAATTAATCTTTTTGCTGATGTCATGTCAGCAGTGTGTGTTGAA
GGCTGTAAAGCAGCCATTTGTTCAGTTTATTTTTCCATTGAACAAGTATTTATCAAAA
ACATACTTTGTGGCAGTCACTATGCTAGGAGCTATGAATACAGAAGGAAAAGTAAA
TGCTCTTGGATACTACACTCCAGTTGTGATAAAAAAGAAAAAATGTATTCTTCACCA
ACTTCAACATCTTGATGTGCAAAAACATAATACATGAATTAGATCTACCTAATTACA
CAGAATTAGACCAATTGTTTCTGGAATTGTGGGCTCATATTTTTAATAACTGTCCTCC

TGCCTCTCTGTCGACAGGTTTTATAAATATTCATTTAATTACACACACACACACGAA
CAATTGACTAGTACTTGCTCTCATTCTTCTAGATGTCATCACAATGGATCCCAAAGA
CAATTGTTCAAAAGATGCAAATGGTAAGCTTTTGTGTTTTTCCCTTCCTCCTGATCAT
TTTGTTTTGAACTTCTCTGGCTTGAAAAATCAGGGAATGGATTTTGCTAGGTTGGATG
CTGCAGAATGGACCTAGTGATATTTTAAATTAGTCCCTCATTTTCTAGGAGTTGTATT
AACAAACCTAACTACTGCTTTGGGGTATGAGATGACTGTAAATTAGAGAGGGTACA
GTGGTATAGTGATATGCTTTTAATTATTTCAAAAAAAAGATTTTATTCATTCATGTGT
CTTTTTTCTTTTTCTTTTCTTTTTTTTTTTTTTTTGGACAGAGTCTTGCTCTGTCACCCA
GGCTGGAGTGCGGTGGCAGTATCTCAGCTCACCACAACCTCCGCCTCCCGGCTTCAA
GTGATTCTCCTGCCTCAGCTTCTCGAGTAGCTGGGACTACAGGCGCGTGCCACCATG
CCCGGCTAATTTTTGTATTTTTAGTAGAGTTGGGGTTTCACCATGTTGGCCAGGATGG
CCTCGAATTTGTGACCTCGTGATCTGCCCCCTCGCCCTCCCGAACTGTTGGGATTACA
GGCGTGAGTCACTGTGCCCGGCCTCCTGTCCTGTCTTTTGTTTAATGACTGGGAAAA
ACATGATACCATGTTGCTTCTCGAGTTGTTTTGTTTTAGTCTTTGGTCTTTGCTAGTAG
CTAATAACACGAACTAGTGTTTATCAAGTGCTTTTTACACAGAAGGGCTTGGGCTGT
GTTCTGCATTTTCTTGTTTAACCCTCTTAAAACTCCTATAAAATGGTACATATTTTTCT
CCCAATTTACAGTCCCTTTAAAGCAAATAATTATAAAAATCCCTATACATGTCACAC
AGCTAGATCTGGGATTTCAAATCAGGCCATCAAACAAAGAGTTTATGTACTTAGTAA
GTTTTCTGTTCTTTTTCTACAATAGAGTCAGATAGCAAGAAATTACCAAGCCAGGAA
CCTGAAACAAAACGGACATCATGTGGGGCTGGGTGGGTGCATGGGCTTTGCAGACT
GGACTTTCACTCCAGCTCTTTTAATGATTAGGTGTAAGTGACCTACATTTTGTGAGCA
ACAGTTTTCTCATCAGCCAACAAAGAATAATTACACCAGATTCACAGTTATTGAAGA
GATAAAGGCATGAATGTGAGATGTCTGGCATAGGGCATCTCATTTAGCAGACACAG
AATGAGTACTTGTTTCTGGCTTTTTCTCTCTACATATGCACAAAGAATGCGACTAGA
AGCATGGGCTCTAGCCCTGCTCAACTTTCCTCTATTTCCAATACCAAGGGGCTCTGA
CTTAGGCTGCCACACCAGGCAAGGAGGGCAGTACCACCTCACTTGACCAAGGGCAG
GGAGTCACGGACACATCACTTCTTGAGATCCTTTTCCACACCAAGGACTGATGTTTC
TGGAATTCTCACTTTATGAAGACAAAACATATAAATGGAAATTTTCTCAGGTAGAGA
CTCACTCTTGTAGCTCATTGAGTAGGCACTAGTGGTCCACCCCCACTGTCTTTACTTA
TTCCTTGACATCACATATCTCTTGCAAAACCTCAAATAATATTAAATGCAATCACCC
AATAATAGCATAGCCATAATTAGAGGCATTTAGGAAAGACAGGTGAGTGTGCCACA
ACTACCTAACACATCAGCAAATCTGGATTAACCACTTTCTTTGATTTTCCACAATGCA
ACCTTACTTTTTAATAGTTGGGAATGTTCTAAGTGAATTTAGCAGAGGTTGTTAATCA
ACTTGAAAGCTGAATTCTGACTTGTCTGACTCTTGGTGGTGCTGGTAGCAGTAGATG
TTTACTTTTAGGTTTTGGTGGTGGTGGAATATCACTTCAACGTAAATCATCAGAAAT
AAGTATTTGTGAACCCCTCTCGCATTAATGTATCTTATTCTGTAAAAAGAACATGTG
CAATTTCTCTTAGATACACTACTGCTGCAGCTCACAAACACCTCTGCATATTACATGT
ACCTCCTCCTGCTCCTCAAGAGTGTGGTCTATTTTGCCATCATCACCTGCTGTCTGCT
TAGAAGAACGGCTTTCTGCTGCAATGGAGAGAAATCATAA (SEQ ID NO: 187)

ATAAACAACTTGATGCAGATGTTTCCCCCAAGCCCACTATTTTTCTTCCTTCGATTGC
TGAAACAAAACTCCAGAAGGCTGGAACATACCTTTGTCTTCTTGAGAAATTTTTCCC
AGATATTATTAAGATACATTGGCAAGAAAAGAAGAGCAACACGATTCTGGGATCCC
AGGAGGGGAACACCATGAAGACTAACGACACATACATGAAATTTAGCTGGTTAACG
GTGCCAGAAGAGTCACTGGACAAAGAACACAGATGTATCGTCAGACATGAGAATAA
TAAAAACGGAATTGATCAAGAAATTATCTTTCCTCCAATAAAGACAGGTATGTGTTT
ACACATATCATCTGTCAGAACACTTCTTTGAAAGTGAATGCTGCATTTTTTCCTTTCA

GTATTAATGAAAAACATAAATCTTTCTTAAAAATTGTTACATTTAATGGTAGCGTAA
ATGCCCTGCTACTTTTCTATAGAATTAAAATGGTATAGGTTTTGGAGAAAACAAAAT
TGAAAAAGTTGCTGAAGGTTTGTCAGCCTCAGCTCCATTATCCAAAATAAGAAAGTC
ACGTGCTGGTTTTTAGGGTTGTTAGATGGATTAAAGAAACAACATACACAGAAGCAT
CTAGCAACGTGACACGTGGTAAACGCTCAAAAAGTGTTCTCCCTTCTTTTGATGACT
TTACTTGATCAGGAAATAACATATATATGTCTTTCAGGAATGTTCTGCCCAAGCAGG
AGAGTCACTCACCTCAATCTTGCTACCCACAAAGTTTAACCTAAAAACAACGGGTTC
ATTGTTGACAAAATAATGTTTATCTGAAGATAACTGTAGATCATATTTATCTGTAGA
TAATGTTTATCTGTGGAGTGTGGCTCTACAAAACATAGAATAGTCTTGGTCACTGCA
GTTTTATAGAGGCCTTGGGTTTTTCAGAGTTTCATTTTATATATCACCATAAAGTAAC
ATTTCATAATTACAGGTTGGTAAGGCTTACATGTACAAACATTCTTCCATTTTCCATA
ATAAATGCATTTCCTGCCATTGGTGAATGCAGCTCAATAAACATTTATTGTACAATT
ATGACACGCCAGGCTTAGTGGAAATGTGGATGAACAGACAAGGATGAGTTACTGTC
CTAAGGATGATGCATGACAGTGCAGAGAATATACTCTCTTCCTGATCACTCAGGGTC
ACTCATGATTCATGCGCGAGGTCCCAAAACAGTGCCTTTGATGCAGATTCTGTACAT
CTCTAGACGATTGGTCCAAGGGCTGAATGTGCTCTGGCCCAGTGGTCCAGTCTGTCA
CTATATGTCAACATCCTGAATATGAACATAACAGTCCAACATCTCAAGAGTGGGCAT
GAAAAGGACTCATTTTGTGCTTTTTCCTGTGGTTAACAAGTCCTTTTTAGCCTGGGGG
AACAAGCATTAACAAAATGTTTGAAGATCTTTGCCACGTACCATTCCAAATTTCTAG
GGTAAGTCTTTAGCTTTTCAGATCCTGAGTTTCTGCAATGATCAAATGTGATTTGGAC
AGTTGCGTTGACTTTCTCCTGGGGCTATAATGGAGTGCAAAGGAAACAATGGCAGG
GAAAATGCTTGCTTTCAAAATGGTAGCATGGATGTGTTCATTCGTGTAGTTACTGTA
TTAGGTATAGCCTTTCCTGAAACTAACTGAAGTGGGGTTATAAAAACAGTCCCAATT
TTCTATTTCCTTTGCTGAGACACAAAGAGGAGACAAAAGAGCAAAGCTTGAGGGTA
GTTTTACCACTGTGCTTAAGTGTTCTGATTTTTCCAGTGATCAGGGTGAAATAAAAA
GCATAGTAAGTTCCAGGGCAGTGAATACCATACAGGAGACAAGTTACAGTTTTATA
ATGTGTTTTACTTTACACTAAATTCTAAAAGTAAAATGTCTTTTTTTTTTTCCGAGAC
AGAGTTTCACTCTTGTAGCCCAGGCAGGAGTGCTATGGTGTGATCTCGGCTCACAGC
AACCTCCACCTCCCAGTTTCAAGCGATTCTTCTGCCTCAGCCTCCCGAGAAGTTGAA
ATTACAGGTGCCTGGCACCATATCTCGCTAATTATTCTATTTTTAGTAGAGATCGGGT
TTTACCATGTTGGCCAGGCTGGTCTCGAACTCCTGACTTCAAGTGATCCACCCGCCT
CAGCCTCCCAAAGTGCTGGGATTACAGGTGTGAGTCACTGTGCCGGACCTAACAGTA
AAATGTCTTTCATGTGCTTCTCAAGGCAACTACATTAAGGAGGACACATCTCTTAAT
GTCATTCTACAGTAGATTTCTAATGCTCTTTCTTGGAAGTTTGTTTTTCTGAGAAGAG
CTAAAAATATAATAACATGGAAGTGATCATATTATATAATCAATGAAGTGCTTTCAA
AGGAGATAAAACTAACCTGGTCTGCATTTGCAACCAGCCTTGATTGAGAGAGAGAG
AACTCAGGATACACTTAGAGATTTTATTATGGGGAATAGTTACTTTATTCATTTTACC
TCAATCAATGCATGGAAATAAGTGACAGTCATTTTCATTTATCTTTTAATAAATAAA
GTCACCATGAGGAAAATGAAAACCCATTAAAGTCAGTCCTTAAAGATATTTGGACA
TGCAGACATGATAACTAACATTTCCATTCGTGAGACTTACCCAAAACCTATACCTCA
AGTCCATTTCTTAGAATACATGAAATAAAGATCTCAGTGAGTGTATAAAACTGCACA
CCAGAATCATATCCGTATAGACAAGAATACATCTACTAGAAAAATATAAACCAAAA
CACCAAGGTGACTCTGTTTTTTTCTGTTTTAAAATATGTTGTCTTTGTATGCATGTTTG
CTTCTTCCTTTTTTTTTTTAAACATCGCAGATAAATTCAACTCTCACCTCAGTTGAGA
GAGAACTGTCAATGTGACTTGGCCTCTCTCTTTCTAGTCCCAGAAAGAATTGCACTG
AAATGCTGAGCTCCTGTAATAAAAATGACCATTTGCTGAGAGTAATTAACATACTGA

AAGAGATTTTCTTAGAATAGTGCACAATGGCCCAATGGTGACATTATATTGTCTCTT
TATAAATTATTTTCTATCTATTTCTGTGGATTATTTCTACAAAGCACTTTTCATATGTC
CAATTCCTTTTATTCCCCTACAAGTACTGACTGACTACTGGCTCTGCTGTTCACTGAT
ATGACTTTCGGCAAGTTGCCTGCACTTTTTAAACGTTATTTCCTCATTCAGAACATGG
GGCCATACAAAATACAACTCACTTCAGTGTTATTGGGGAATTAAACAAATAAATGC
ATGGGAAGCATTTAACATAGTGCCTGACACAATAATGAGCACTCAGTAGATGTTAG
CTTTTATTAATATTGTTGTTGCTATGTCCAGAAACACTATACCTCCAGAAAATCATGG
GTACTTGCTGGGGACGTTGGGGATATGCATGATTTTGAAAGGAGTGACTGCTCTTTA
CTGCTCAGATGAGAAATTTTTCTAAGCCAGACTCCTTCAAACATGTAAGATTCTGTT
GTGGATTCTAGGACTGAAAGAATTCTTGGCCGAGTGTGGTGGCTTATCCTGGTAATC
TCATCATTTGGGAGGACAAGGCAGGAAGATTGCTTGAGCCCAGGAGTTGGAAACAA
GCCTGGACAACATGGCGAAACCCTGTCTCTACAAAAAATACAAACATTAGCTGGTC
ATGGGAGTGAGTGCCTGTACTCCCAGCTACTCAGGAGGCTAAGATAGGAGGATCAC
CTGAGCCTGGGCAGTTTGAGGTTTCAGTGAGCCGTGATGACACCATACTATACTCCA
CTCCAGCCTGGGTGACAGTGACATCCTGCCTCAAAAAAACCCCCAAAATTATTCTTT
TTGCTGATTTCATGTCAGCAGTGTGTGCTGAAGGCTGTAAAGTAGCCACTTGTTCTGT
TTATTTTTCCATTGAACAAGTATTTATCAAAAACGTACTTTGTGGAAGGCACTGTGCT
AGGAACTATGCATACAGAAGGAAAACCAAATGTTCTTGGATACTACACTCCAGTTGT
GATAAAAAAGAAAAAAGTATTCTTCACAAACTTCAACATTTTGATGTGCAAAAACA
TAATATATGAATTAGATCTACCTAACTACACAGAATTAGACCAATTATTTCTGGGAT
TATGGGCTCATATTTTTAATAACTGTCCTCCTACCTCTCTGTTGACAGGTTTTATAAA
TATTCATTTAATTACACACAGTCACAGACACACTCAGACACACACACATACACACAC
ACACACACCTTGACAAATAATGGGCATGAACAATTGACTGGTACTTGCTCTCATTCT
TCTAGATGTCACCACAGTGGATCCCAAATACAATTATTCAAAGGATGCAAATGGTAA
GTTTTTGTGTTTTTTATTTCCTCCTGATCATTTTAAGTTTTGAACTTCTCTGGCTTGAA
AAATCAGGGAATGGATTTTGCTAGGTTGGATGCTGCAGAATGGACCTAATCATATTT
TAAATTAGTCCCTCTTTTTCTAGGAGTTGTATTAACAAACCTAACTACTGCTTCATGT
AAGAGATGACTGTAAATTGAAGGGTACAGTGATATGCTTTCAGTTATTTCAAAAAAC
AGACTTTACTCATCCATGTGTCTTTTTTCTTTTCTTTTTTTTCTTTTTTGAGACGGAGT
CTCGCTCTGTTGAACAGGCTGGATTGCAGTGACGCGATCTCACCTCACTACAACCTC
CGCCTCTGGAGTTCAAGCGATTCTCCAGCCTCAGCTTCTCAAGTAGCTGGGACTACA
GGCACATGCCACCATGTCCGGGTCATCTTTGTATTTTTAGCAGAGACCGGGTTTCAC
TATGTTGGCCAGGCTGGTCTAGAATTCCTGACTTCGTGATCTGCCCCCTCAGCCCTCC
GAAGTGCTGGGATTACAGACGTGAGTCACTGTGCCCGGCCTAACAGTAAAATGTCTT
TCATGCGCTTCTCAAGGCAACTACGTTAAGGAGGACACTTCTCTTAATGTCATTCTA
CAGTAGATTTCTAATGCTCTTTCTTGGAAGTTTGTTTTTCTGAGAAAAGCTAAAAATA
TAACATGGAAGTGATCATATTGTATAATCAATGAAGTGCTTTTCAAGGAGATAAAAC
TAATCTGGTCCACGTTTGCAACCAACCTTGATTGAGAGAGAGAGAGAACTCAGGAT
ACACTTGGAGATTTTATTATGGGGAATAGTTACTTTATTCTTTTTTCCTCAATCAATT
CATGGAAATAAGTGATAGTCATATTCATTTATCTTTTAATAAATGAAGTCACCATGA
GGAAAATAAAAAGACATTGAAAACCCATTAAAGTTAGCCCTTAAAGATATTTGGAC
ATGCAGACTTGATAACTAACGTTTGCATTCTTGAGACTTACCCAAAACCCATACCTC
AAGTCCATGTTTTTAGAATTCATGAAATAAAGATCTCAGTGAGTGCATAAAATTGCG
CACCAGAATCATATCCGTATAGACAAGAACACATCTACTAGAAAAATAATAAACCA
ACACACCAATGCAACTGTGTTTTCTTCTGTTTTAAAATATGTTGTCTTTGTATGCATG
TTTGCTTCTTCCTTTTTTTTTTTTAACATCACAGATAAATTCAACTCTCACCTCAGGTT

TTATTGAGAGAACTGTCAATGTGACTTGGCCTCTGTCTTTCTAGTCCCAGAAAGAAT
CGCACTGAAATGCTGAGCTCCTGTAATAAAAATGACCATTTGCTGAGAGTAATTAAC
ATACTGAAAGAGATTTTCTTAGAGTACACAATGGTGACATTATATTGTCTCTTTATA
AATAACTTTCTATCTATTTCTGTGGATTATTCCTACAAAGTACTTTTCATATGTCCAG
TTTCTTTTCTTCCCCTACAACTACCGTCTGAATACTGGCTCTGCTATTTGCTGATATG
ATTCTCGGCAAGTTGCCTGCACTTTTTAAACTTTATTTCCTCATTCAGAACATGGGGC
CATGTAATACTCATGTACGTGAGTATTACGTAATAATGCTCACTTAAGTGTTACTGG
GGAATTAAACAAAAAAATGCATGGCAAGCATTTAACATAGTGCCTGACACAATAAT
GAGCACTCAGTAGATGTTAGATTTTATTAATATTGTTGTTGTTATGTCCGGAAACACT
ATACCTCCAGAAAATCATGGGTACTTGCTTGGGATGTTGGGGATATGCATGATTTGG
AAAGGTATGACTGCTTTTTTCTGCTTAGATGAGAAATTTTTCTAAGCCAGACTCCTTC
AAATATGTAAGATTCTGTTGTGGATTCTAGGACGGAAAGAATTCTTGGTCAGGTGTG
GTTTCTTATCCCTGTAATCCCAGAATTTTGGGAGGACAAGGCAGGAAGATTGCTTGA
GCCCAGGAGTTTGAAACCAGCCTGGGCAACAAGACGAAACCCTGTCTCTACAAAAG
TACATAAATTAGCTTGGCTTGGTGGTGTGTGCCTGTATTACCAGCTATTCGGGAGAC
TGAGATGGGAGGATCTCCTGAACCTGTGAAGTTTGAGGCTTCAGTGAGCCGTGATGA
CACCATACTATACTCGACTCCAGCCTGTGCGACAGTGAGACTCTGCGTCAAAAAAA
AAACCCCAAAATTATTGTTTTTGCTGATTTCAGGTCAGCAGTGTGTGCTGAAGGGTG
TAAAGTAGCCACTTGATCAGTTTATTTTTCCACTGAACAAGTATTTATCAAAAACAT
ACTTTGTGGTCTGTTTTTGATAAATAAAAAGGCACTGTGCTAGGAGCCATGAATACA
GAAGGAAAACCAAATGTTCTTGGATACTACACTCCAGTTGTGATAAAAAAGAAAAA
TGTATTCTTCACGAACTTCAACATTTTGATATGCAAAAACATAGTATATAAATTAGA
TCTACCTGATTACGTAGAATCAGACCAATTATTTCTGGAATTGAGGGCTCATATTTTT
AATAACTGTCCTCCTGCCTCTCTGTTGACAGGTTTTATAAATATTCATTTAATTACAC
ACACACACACACACACCTTGACAAATAATGGACATGAACAATTGACTAGTACTTGCT
CTCATTCTTCTAGATGTCATCACAATGGATCCCAAAGACAATTGGTCAAAAGATGCA
AATGGTAAGCTTTTGTGTTTTTCCTTTCCTCCTGATCATTTTAAGTTTTGAACTTCTCT
GGCTTGAAAAATCAGGGAATGGGCCGGGTGCGGTGGCTCACGCCTGTAATCCCAGC
ACTTTGGGAGGCCGAGGCGGGCGGATCACGAGGTCAGGAGATCGAGACCATCCCGG
CTAAAACGGTGAAACCCCGTCTCTACTAAAAATACAAAAAATTAGCCGGGCTTAGT
GGCGGGCGCCTGTAGTCCCAGCTACTTGGGAGGCTGAGGCAGGAGAATGGCGTGAA
CCCGGGAGGCGGAGCTTGCAGTGAGCCGAGATTGCGCCACTGCACTCCACTCCAGC
CTGGGCGACAGAGCGAGACTCCGTCTCAAAAAAAAAAAAAAAAAAAAAAAAGAA
AAATCAGGGAATGGATTTTGCTAGGTTGGATGCTGCAGAATGGACCTAGTGATATTT
TAAATTAGTCCCTCTTTTTCTAGGAGTTGTATTAACAAACCTAACTACTGCTTCGGGT
ATGAGATGACTGTAAATTAGAGGGTACAGTGATATGCTTTCAGTTATTTCAAAAAAC
AGACTTTATTCATCCGTCTGTCTTTTTTTTTTTTTTTTTTTTTTTTTTGAGACGGAGG
AGTCTCACTCTATCACCCAGGCTGGAGTGCAGTGGCGCGATCTCGGCTCACCATAAC
CTCCGCCTTACTGGTTCAAGCGATTCTCCAGCCTCAGCTTCTCAAGTAGCTGGGACT
ACAGGTGCACACCACCATACCTGGCTAATTTTTGTATTTTTAATAGAGATGGGGTTT
CACCACGCTGGCCAGGATGGTCTTGAATTCTTGACCTCGTGATCTGCCCCCTCGGGC
TCCCAAACTTCTGGGATTATAGGCGTGAGCCACTGTGCCCGGCCTTCTGTCTTTTGTT
ATAATGACTGGGGAAAACATGATACCATGTTGCTTCTTGAGTTGTTTTGTTTTAGTCT
TTGGTCTTTGCTAGTAGCTAATAACACGAACTAGTGTTTATCAAGTGCTTTTTACACA
GAAGGGCTTGTTCTGCATTTTCTAGTTTAATCATCTTAATACTCCTATAAAGTAGTAC
AATATATTTTCTCCCATTTTACAGTCCCTTTAAAGTAAATAACTATAAAAATCCCTTA

TACATGTCACACAGCTAGGTCTGGCATTTCAAATCAGGACATCAAACAAAGAATTCG
TGCAGTTACTAAGTCCTCTATTTTTTCTACAATAGAAAAAATAGCAAGAATTACAGA
TAGCAAGACATTACAAGGCAGGAATCTGAAACGAAAGGGACATAATGTGGGGCTGG
GTGGGTGCATGAGCTTTGCAGACTAGACTTTCATTCCAGCTCTTTTAATGATTAGGTG
TAAGTGACCTACATTTTGTGAGTAACAGTTTTCTCATCAGCCAACTAAGAATAATTA
CACCAGATTCACAGTTATTGAAGAGATAAGGGCATGAATGTGAGATGTCTGGCGTA
GGGTATCTCATTTAGCAGACACAGAATGAATACTTGTTTCTGGCTTTTTCTCTCTACA
TATGCACAAAGAATGTGACTAGAAGCATTGGCTCTAGCCCTGCTCAACTTTCCTCTA
TTTCCAATACCAAGGGGCTCTGACTTAGGCTGCCACACCAGGCAAGGAGGGGCAGT
ACCACCTCACTTGACCAAGGGCAGGGAGTCACGGACACATCACTTCCTGAGATCCTT
TTCCACACCAAGGACTGATGTTTCTGGAATTCTCACTTTATGAAGACAAAACATATA
AATGGAAATTTCTGCAGGAAGAGACTCACTCTTGTAGCTCATTGAGTAGGCACTAGT
GGTCCACCCCCACTGTCTTTACTTATTCCTTGACATCACATATCTCTTGTAAAACCTC
AAATAATGTTAAATGCAATCACCCAATAATAGCATAGCCATAATTAGAGGCATTTAG
GAAAGACAGGTGAGTGTGCCACAACTACCTAACACATCAGCAAATCTGGATTAACC
ACTTTCTTTGATTTTCCACAATGCAACCTTACTTTTTAATAGTTGGGAATGTTCTAAG
TGAATTTAGCAGAGGTTGTTAATCAACTTGAAAGCTGAATTCTGACTTGTCTGACTC
TTGGTGGTGCTGGTAGCAGTAGATGTTTACTTTTAGGTTTTGGTGGTGGTGGAATATC
ACTTCAACGTAAATCATCAGAAATAAGTATTTGTGAACCCCTCTCGCATTAATATAT
CTTATTCTGTAAAAAGAACATGTGCAATTTCTCTTAGATACACTACTGCTGCAGCTC
ACAAACACCTCTGCATATTACACGTACCTCCTCCTGCTCCTCAAGAGTGTGGTCTATT
TTGCCATCATCACCTGCTGTCTGCTTAGAAGAACGGCTTTCTGCTGCAATGGAGAGA
AATCATAA (SEQ ID NO: 188)

**[0395]** In some embodiments, the constant domain of a presently disclosed HI-TCR comprises a native or modified TRDC peptide. In some embodiments, the TRDC polypeptide has an amino acid sequence that is at least about 85%, about 90%, about 95%, about 96%, about 97%, about 98%, about 99% or 100% homologous or identical to the amino acid sequence set forth in SEQ ID NO: 189, which is provided below.

SQPHTKPSVFVMKNGTNVACLVKEFYPKDIRINLVSSKKITEFDPAIVISPSGKYNAVKLG
KYEDSNSVTCSVQHDNKTVHSTDFEVKTDSTDHVKPKETENTKQPSKSCHKPKAIVHTE
KVNMMSLTVLGLRMLFAKTVAVNFLLTAKLFFL(SEQ ID NO: 189)

**[0396]** In some non-limiting embodiments, the HIT constant region comprises a hinge/spacer region that links the extracellular antigen-binding domain to the constant domain. The hinge/spacer region can be flexible enough to allow the antigen binding domain to orient in different directions to facilitate antigen recognition. In some non-limiting embodiments, the hinge/spacer region can be the hinge region from IgGl, or the CH2CH3 region of immunoglobulin and portions of CD3, a portion of a CD28 polypeptide, a portion of a CD8 polypeptide, a variation of any of the foregoing which is at least about 80%, at least about 85%, at least about 90%, or at least about 95% homologous or identical thereto, or a synthetic spacer sequence. In certain non-limiting

embodiments, the hinge/spacer region of the CAR can comprise a native or modified hinge region of a Oϋ3z polypeptide, a CD40 polypeptide, a 4-1BB polypeptide, an 0X40 polypeptide, a CD 166 peptide, a CD 166 peptide, a CD8a peptide, a CD8b peptide, an ICOS polypeptide, an ICAM-l peptide, a CTLA-4 peptide, a synthetic peptide (not based on a protein associated with the immune response), or a combination thereof.

**[0397]** In some non-limiting embodiments, the HIT comprises an antigen binding chain which does not comprise an intracellular domain. In some embodiments, the antigen binding chain is capable of associating with a CD3ζ polypeptide. CD3ζ polypeptides are discussed in detail above. In some embodiments, the antigen binding chain comprises a constant domain, which is capable of associating with a CD3ζ polypeptide. In some embodiments, the CD3ζ polypeptide is endogenous. In some embodiments, the CD3ζ polypeptide is exogenous. In some embodiments, binding of the antigen binding chain to an antigen is capable of activating the CD3ζ polypeptide associated to the antigen binding chain. In some embodiments, the exogenous CD3ζ polypeptide is covalently-linked to or integrated with a costimulatory molecule disclosed herein.

**[0398]** In some non-limiting embodiments, the HIT comprises an antigen binding chain that comprises an intracellular domain. In some embodiments, the intracellular domain comprises a CD3ζ polypeptide. In some embodiments, binding of the antigen binding chain to an antigen is capable of activating the CD3ζ polypeptide of the antigen binding chain.

**[0399]** The activated CD3ζ polypeptide can activate and/or stimulate an immunoresponsive cell (e.g., a cell of the lymphoid lineage, e.g., a T cell). In some embodiments, the CD3ζ polypeptide comprises three immunoreceptor tyrosine-based activation motifs (ITAM1, ITAM2 and ITAM3), three basic-rich stretch (BRS) regions (BRS1, BRS2 and BRS3), and transmits an activation signal to the cell (e.g., a cell of the lymphoid lineage, e.g., a T cell) after antigen is bound to the antigen binding chain. The intracellular signaling domain of the CD3ζ -chain is the primary transmitter of signals from endogenous TCRs. In some embodiments, the CD3ζ polypeptide comprises or has an amino acid sequence that is at least about 85%, about 90%, about 95%, about 96%, about 97%, about 98%, about 99% or about 100% homologous to the sequence having a NCBI Reference No: NP_932170 (SEQ ID NO: 190, NCBI Reference No: NP_000725.1 (SEQ ID NO: 191) or fragments thereof, and/or may optionally comprise up to one or up to two or up to three conservative amino acid substitutions. In some non-limiting embodiments, the CD3ζ polypeptide

comprises or has an amino acid sequence that is a consecutive portion of SEQ ID NO: 228, which is at least 20, or at least 30, or at least 40, or at least 50, and up to 164 amino acids in length. Alternatively or additionally, in non-limiting various embodiments, the CD3ζ polypeptide comprises or has an amino acid sequence of amino acids 1 to 164, 1 to 50, 50 to 100, 100 to 150, or 150 to 164 of SEQ ID NO: 190. In some embodiments, the CD3ζ polypeptide comprises or has an amino acid sequence of amino acids 52 to 164 of SEQ ID NO: 190.

MKWKALFTAAILQAQLPITEAQSFGLLDPKLCYLLDGILFIYGVILTALFLRVKFSRSADA PAYQQGQNQLYNELNLGRREEYDVLDKRRGRDPEMGGKPQRRKNPQEGLYNELQKDK MAEAYSEIGMKGERRRGKGHDGLYQGLSTATKDTYDALHMQALPPR (SEQ ID NO: 190)

In some embodiments, the intracellular signaling domain comprises a human CD3ζ polypeptide. The human CD3ζ polypeptide can comprise or have an amino acid sequence that is at least about 85%, about 90%, about 95%, about 96%, about 97%, about 98%, about 99% or about 100% homologous or identical to SEQ ID NO: 191 or fragments thereof, and/or may optionally comprise up to one or up to two or up to three conservative amino acid substitutions. SEQ ID NO: 191 is provided below:

RVKFSRSADAPAYQQGQNQLYNELNLGRREEYDVLDKRRGRDPEMGGKPRRKNPQEG LYNELQKDKMAEAYSEIGMKGERRRGKGHDGLYQGLSTATKDTYDALHMQALPPR (SEQ ID NO: 191)

An exemplary nucleic acid sequence encoding the amino acid sequence of SEQ ID NO: 191 is set forth in SEQ ID NO: 192, which is provided below.

AGAGTGAAGTTCAGCAGGAGCGCAGACGCCCCCGCGTACCAGCAGGGCCAGAACC AGCTCTATAACGAGCTCAATCTAGGACGAAGAGAGGAGTACGATGTTTTGGACAAG AGACGTGGCCGGGACCCTGAGATGGGGGGAAAGCCGAGAAGGAAGAACCCTCAGG AAGGCCTGTACAATGAACTGCAGAAAGATAAGATGGCGGAGGCCTACAGTGAGATT GGGATGAAAGGCGAGCGCCGGAGGGGCAAGGGGCACGATGGCCTTTACCAGGGTCT CAGTACAGCCACCAAGGACACCTACGACGCCCTTCACATGCAGGCCCTGCCCCCTCG C (SEQ ID NO: 192)

**[0400]** In some non-limiting embodiments, the HIT antigen binding chain comprises an intracellular domain, wherein the intracellular domain comprises a co-stimulatory region. In some embodiments, the intracellular domain comprises a co-stimulatory region and a CD3ζ polypeptide. In some embodiments, the intracellular domain comprises a co-stimulatory region and does not comprise a CD3ζ polypeptide. In some embodiments, the co-stimulatory region comprises at least one co- stimulatory molecule, which can provide optimal lymphocyte activation. As used herein,

"co-stimulatory molecules" refer to cell surface molecules other than antigen receptors or their ligands that are required for an efficient response of lymphocytes to antigen. The at least one co-stimulatory signaling region can include a CD28 polypeptide, a 4-1BB polypeptide, an 0X40 polypeptide, an ICOS polypeptide, a DAP-10 polypeptide, or a combination thereof. The co-stimulatory molecule can bind to a co-stimulatory ligand, which is a protein expressed on cell surface that upon binding to its receptor produces a co-stimulatory response, i.e., an intracellular response that effects the stimulation provided when an antigen binds to its CAR molecule. Co-stimulatory ligands, include, but are not limited to CD80, CD86, CD70, OX40L, and 4-1BBL. As one example, a 4-1BB ligand (i.e., 4-1BBL) may bind to 4-1BB (also known as "CD137") for providing an intracellular signal that in combination with a CAR signal induces an effector cell function of the CAR+ T cell. CARs comprising an intracellular signaling domain that comprises a co-stimulatory signaling region comprising 4-1BB, ICOS or DAP-10 are disclosed in U.S. Patent No. 7,446,190.

**[0401]** In some embodiments, the co-stimulatory region comprises a co-stimulatory signaling region that comprises a CD28 polypeptide. The CD28 polypeptide can comprise or have an amino acid sequence that is at least about 85%, about 90%, about 95%, about 96%, about 97%, about 98%, about 99% or 100% homologous or identical to the sequence having a NCBI Reference No: P10747 or NP_006130 (SEQ ID NO: 193), or fragments thereof, and/or may optionally comprise up to one or up to two or up to three conservative amino acid substitutions. In non-limiting embodiments, the CD28 polypeptide comprises or has an amino acid sequence that is a consecutive portion of SEQ ID NO: 193 which is at least 20, or at least 30, or at least 40, or at least 50, and up to 220 amino acids in length. Alternatively or additionally, in non-limiting various embodiments, the CD28 polypeptide comprises or has an amino acid sequence of amino acids 1 to 220, 1 to 50, 50 to 100, 100 to 150, 114 to 220, 150 to 200, or 200 to 220 of SEQ ID NO: 193. In some embodiments, the co-stimulatory region comprises a co-stimulatory signaling region that comprises a CD28 polypeptide comprising or having an amino acid sequence of amino acids 180 to 220 of SEQ ID NO: 193.

MLRLLLALNLFPSIQVTGNKILVKQSPMLVAYDNAWLSCKYSYNLFSREFRASLHKGLD
SAVEVCWYGNYSQQLQVYSKTGFNCDGKLGNESVTFYLQNLYWQTDIYFCKIEVMYPP

PYLDNEKSNGTIIHVKGKHLCPSPLFPGPSKPFWVLVWGGVLACYSLLVTVAFIIFWVRS KRSRLLHSDYMNMTPRRPGPTRKHYQPYAPPRDFAAYRS (SEQ ID NO: 193)

**[0402]** In some embodiments, the co-stimulatory region comprises a human intracellular signaling domain of CD28. The human intracellular signaling domain of CD28 can comprise or have an amino acid sequence that is at least about 85%, about 90%, about 95%, about 96%, about 97%, about 98%, about 99% or about 100% homologous or identical to SEQ ID NO: 194 or fragments thereof, and/or may optionally comprise up to one or up to two or up to three conservative amino acid substitutions. SEQ ID NO: 194 is provided below:

RSKRSRLLHSDYMNMTPRRPGPTRKHYQPYAPPRDFAAYRS (SEQ ID NO: 194).

**[0403]** An exemplary nucleic acid sequence encoding the amino acid sequence of SEQ ID NO: 232 is set forth in SEQ ID NO: 195, which is provided below.

AGGAGTAAGAGGAGCAGGCTCCTGCACAGTGACTACATGAACATGACTCCCCGCCG CCCCGGGCCCACCCGCAAGCATTACCAGCCCTATGCCCCACCACGCGACTTCGCAGC CTATCGCTCC (SEQ ID NO: 195)

**[0404]** In some embodiments, the co-stimulatory region comprises a co-stimulatory signaling region that comprises two co-stimulatory molecules, e.g., co-stimulatory signaling regions of CD28 and 4-1BB or co-stimulatory signaling regions of CD28 and OX40.

**[0405]** 4-1BB can act as a tumor necrosis factor (TNF) ligand and have stimulatory activity. The 4-1BB polypeptide can comprise or have an amino acid sequence that is at least about 85%, about 90%, about 95%, about 96%, about 97%, about 98%, about 99% or about 100% homologous or identical to the sequence having a NCBI Reference No: P41273 or NP_001552 (SEQ ID NO: 196) or fragments thereof, and/or may optionally comprise up to one or up to two or up to three conservative amino acid substitutions.

MGNSCYNIVATLLLVLNFERTRSLQDPCSNCPAGTFCDNNRNQICSPCPPNSFSSAGGQR TCDICRQCKGVFRTRKECSSTSNAECDCTPGFHCLGAGCSMCEQDCKQGQELTKKGCK DCCFGTFNDQKRGICRPWTNCSLDGKSVLVNGTKERDWCGPSPADLSPGASSVTPPAPA REPGHSPQIISFFLALTSTALLFLLFFLTLRFSWKRGRKKLLYIFKQPFMRPVQTTQEEDG CSCRFPEEEEGGCEL(SEQ ID NO: 196)

**[0406]** In some embodiments, the co-stimulatory region comprises an intracellular signaling domain of 4-1BB. The intracellular signaling domain of 4-1BB can comprise or have an amino acid sequence that is at least about 85%, about 90%, about 95%, about 96%, about 97%, about

98%, about 99% or about 100% homologous or identical to SEQ ID NO: 197 or fragments thereof, and/or may optionally comprise up to one or up to two or up to three conservative amino acid substitutions.

KRGRKKLLYIFKQPFMRPVQTTQEEDGCSCRFPEEEEGGCEL (SEQ ID NO: 197) .

**[0407]** An exemplary nucleic acid sequence encoding the amino acid sequence of SEQ ID NO: 197 is set forth in SEQ ID NO: 198, which is provided below.

AAACGGGGCAGAAAGAAACTCCTGTATATATTCAAACAACCATTTATGAGACCAGT ACAAACTACTCAAGAGGAAGATGGCTGTAGCTGCCGATTTCCAGAAGAAGAAGAAG GAGGATGTGAACTG (SEQ ID NO: 198)

**[0408]** An OX40 polypeptide can comprise or have an amino acid sequence that is at least about 85%, about 90%, about 95%, about 96%, about 97%, about 98%, about 99% or about 100% homologous or identical to the sequence having a NCBI Reference No: P43489 or NP 003318 (SEQ ID NO: 199), or fragments thereof, and/or may optionally comprise up to one or up to two or up to three conservative amino acid substitutions.

MCVGARRLGRGPCAALLLLGLGLSTVTGLHCVGDTYPSNDRCCHECRPGNGMVSRCSR SQNTVCRPCGPGFYNDWSSKPCKPCTWCNLRSGSERKQLCTATQDTVCRCRAGTQPLD SYKPGVDCAPCPPGHFSPGDNQACKPWTNCTLAGKHTLQPASNSSDAICEDRDPPATQP QETQGPPARPITVQPTEAWPRTSQGPSTRPVEVPGGRAVAAILGLGLVLGLLGPLAILLA LYLLRRDQRLPPDAHKPPGGGSFRTPIQEEQADAHSTLAKI (SEQ ID NO: 199)

**[0409]** An ICOS polypeptide can comprise or have an amino acid sequence that is at least about 85%, about 90%, about 95%, about 96%, about 97%, about 98%, about 99% or about 100% homologous or identical to the sequence having a NCBI Reference No: NP 036224 (SEQ ID NO: 200) or fragments thereof, and/or may optionally comprise up to one or up to two or up to three conservative amino acid substitutions.

MKSGLWYFFLFCLRIKVLTGEINGSANYEMFIFHNGGVQILCKYPDIVQQFKMQLLKGG QILCDLTKTKGSGNTVSIKSLKFCHSQLSNNSVSFFLYNLDHSHANYYFCNLSIFDPPPFK VTLTGGYLHIYESQLCCQLKFWLPIGCAAFVWCILGCILICWLTKKKYSSSVHDPNGEY MFMRAVNTAKKSRLTDVTL (SEQ ID NO: 200)

**[0410]** In some embodiments, mutation sites and/or junction between domains/motifs/regions of the CAR derived from different proteins are de-immunized. Immunogenicity of junctions between different CAR moieties can be predicted using NetMHC 4.0 Server. For each peptide containing at least 1 aa from next moiety, binding affinity to HLA A, B and C, for all alleles, can be predicted.

A score of immunogenicity of each peptide can be assigned for each peptide. Immunogenicity score can be calculated using the formula Immunogenicity score = $[(50\text{-binding affinity})*\text{HLA frequency}]_n$ ; n is the number of prediction for each peptide.

**[0411]** In some embodiments, the HIT is capable of associating with a CD3 complex (also known as "T cell coreceptor"). In some embodiments, the HIT and the CD3 complex form an antigen recognizing receptor complex similar to a native TCR/CD3 complex. In some embodiments, the CD3 complex is endogenous. In some embodiments, the CD3 complex is exogenous. In some embodiments, the presently disclosed HI-TCR replaces a native and/or an endogenous TCR in the CD3/TCR complex. In some embodiments, the CD3 complex comprises a CD3γ chain, a CD3δ chain, and two CD3ε chains.

**[0412]** In some embodiments, the CD3γ chain comprises or has an amino acid sequence that is at least about 85%, about 90%, about 95%, about 96%, about 97%, about 98%, about 99% or about 100% homologous or identical to NCBI reference number: NP_000064.1 (SEQ ID NO: 201, which is provided below) or fragments thereof, and/or may optionally comprise up to one or up to two or up to three conservative amino acid substitutions.

MEQGKGLAVLILAIILLQGTLAQSIKGNHLVKVYDYQEDGSVLLTCDAEAKNITWFKDG
KMIGFLTEDKKKWNLGSNAKDPRGMYQCKGSQNKSKPLQVYYRMCQNCIELNAATIS
GFLFAEIVSIFVLAVGVYFIAGQDGVRQSRASDKQTLLPNDQLYQPLKDREDDQYSHLQ
GNQLRRN (SEQ ID NO: 201)

**[0413]** In some embodiments, the CD3δ chain comprises or has an amino acid sequence that is at least about 85%, about 90%, about 95%, about 96%, about 97%, about 98%, about 99% or about 100% homologous or identical to NCBI reference numbers: NP_000723.1 (SEQ ID NO: 202, which is provided below), NP_001035741. 1 (SEQ ID NO: 203, which is provided below) or fragments thereof, and/or may optionally comprise up to one or up to two or up to three conservative amino acid substitutions.

MEHSTFLSGLVLATLLSQVSPFKIPIEELEDRVFVNCNTSITWVEGTVGTLLSDITRLDLG
KRILDPRGIYRCNGTDIYKDKESTVQVHYRMCQSCVELDPATVAGIIVTDVIATLLLALG
VFCFAGHETGRLSGAADTQALLRNDQVYQPLRDRDDAQYSHLGGNWARNK
(SEQ ID NO: 202)

MEHSTFLSGLVLATLLSQVSPFKIPIEELEDRVFVNCNTSITWVEGTVGTLLSDITRLDLG
KRILDPRGIYRCNGTDIYKDKESTVQVHYRTADTQALLRNDQVYQPLRDRDDAQYSHL
GGNWARNK (SEQ ID NO: 203)

**[0414]** In some embodiments, the CD3ε chain comprises or has an amino acid sequence that is at least about 85%, about 90%, about 95%, about 96%, about 97%, about 98%, about 99% or about 100% homologous or identical to NCBI reference number: NP_000724.1 (SEQ ID NO: 204, which is provided below) or fragments thereof, and/or may optionally comprise up to one or up to two or up to three conservative amino acid substitutions.

MQSGTHWRVLGLCLLSVGVWGQDGNEEMGGITQTPYKVSISGTTVILTCPQYPGSEILW
QHNDKNIGGDEDDKNIGSDEDHLSLKEFSELEQSGYYVCYPRGSKPEDANFYLYLRARV
CENCMEMDVMSVATIVIVDICITGGLLLLVYYWSKNRKAKAKPVTRGAGAGGRQRGQ
NKERPPPVPNPDYEPIRKGQRDLYSGLNQRRI (SEQ ID NO: 204)

**[0415]** In some embodiments, the HIT comprises two antigen binding chains, e.g., VL-TRAC and VH-TRBC, which are capable of dimerizing, wherein the HI-TCR binds to CD19 (e.g., human CD19).

**[0416]** In some embodiments, the HIT comprises an antigen binding chain that comprises an extracellular antigen-binding domain of a VL domain of an antibody and a constant domain of TRAC. In some embodiments, the antibody binds to CD19 (e.g., human CD19). In some embodiments, the antigen binding chain is designated as "VL-TRAC."

**[0417]** In some embodiments, the HIT comprises an antigen binding chain that comprises an extracellular antigen-binding domain of a VH domain of an antibody and a constant domain of TRBC. In some embodiments, the antibody binds to CD19 (e.g., human CD19). In some embodiments, the antigen binding chain is designated as "VH-TRBC."

### b. Exemplary Donor Template Polynucleotides

**[0418]** In some embodiments, the donor template polynucleotide is an HDRT polynucleotide, which can have homology ends to target any desired target locus. In some embodiments, the HDRT polynucleotide comprises a first (i.e., left or 5') homology sequence as set forth in SEQ ID NO: 205.

GATGTAAGGAGCTGCTGTGACTTGCTCAAGGCCTTATATCGAGTAAACGGTAGTGCT
GGGGCTTAGACGCAGGTGTTCTGATTTATAGTTCAAAACCTCTATCAATGAGAGAGC

AATCTCCTGGTAATGTGATAGATTTCCCAACTTAATGCCAACATACCATAAACCTCC
CATTCTGCTAATGCCCAGCCTAAGTTGGGGAGACCACTCCAGATTCCAAGATGTACA
GTTTGCTTTGCTGGGCCTTTTTCCCATGCCTGCCTTTACTCTGCCAGAGTTATATTGCT
GGGGTTTTGAAGAAGATCCTATTAAATAAAAGAATAAGCAGTATTATTAAGTAGCC
CTGCATTTCAGGTTTCCTTGAGTGGCAGGCCAGGCCTGGCCGTGAACGTTCACTGAA
ATCATGGCCTCTTGGCCAAGATTGATAGCTTGTGCCTGTCCCTGAGTCCCAGTCCATC
ACGAGCAGCTGGTTTCTAAGATGCTATTTCCCGTATAAAGCATGAGACCGTGACTTG
CCAGCCCCACAGAGCCCCGCCCTTGTCCATCACTGGCATCTGGACTCCAGCCTGGGT
TGGGGCAAAGAGGGAAATGAGATCATGTCCTAACCCTGGAATTGGATCCTCTTGTCT
TACAGAT (SEQ ID NO: 205)

**[0419]** In some embodiments, the HDRT polynucleotide comprises a second (i.e., right or 3') homology sequence as set forth in SEQ ID NO: 206.

AATATCCAGAACCCTGACCCTGCCGTGTACCAGCTGAGAGACTCTAAATCCAGTGAC
AAGTCTGTCTGCCTATTCACCGATTTTGATTCTCAAACAAATGTGTCACAAAGTAAG
GATTCTGATGTGTATATCACAGACAAAACTGTGCTAGACATGAGGTCTATGGACTTC
AAGAGCAACAGTGCTGTGGCCTGGAGCAACAAATCTGACTTTGCATGTGCAAACGC
CTTCAACAACAGCATTATTCCAGAAGACACCTTCTTCCCCAGCCCAGGTAAGGGCAG
CTTTGGTGCCTTCGCAGGCTGTTTCCTTGCTTCAGGAATGGCCAGGTTCTGCCCAGAG
CTCTGGTCAATGATGTCTAAAACTCCTCTGATTGGTGGTCTCGGCCTTATCCATTGCC
ACCAAAACCCTCTTTTTACTAAGAAACAGTGAGCCTTGTTCTGGCAGTCCAGAGAAT
GACACGGGAAAAAAGCAGATGAAGAGAAGGTGGCAGGAGAGGGCACGTGGCCCAG
CCTCAGTCTCTCCAACTGAGTTCCTGCCTGCCTGCCTTTGCTCAGACTGTTTGCCCCT
TACTGCTCTTCTAGGCCTCATTCTAAGCCCCTTCTCCAAGTTG (SEQ ID NO: 206)

**[0420]** In some embodiments, the HDRT polynucleotide comprises a coding sequence for a first (i.e., left or N-terminal) P2A peptide sequence as set forth in SEQ ID NO: 207.

GGATCTGGAGCAACAAACTTCTCACTACTCAAACAAGCAGGTGACGTGGAGGAGAA
TCCCGGCCCC (SEQ ID NO: 207)

**[0421]** In some embodiments, the HDRT polynucleotide comprises a coding sequence for a second (i.e., middle) P2A peptide sequence as set forth in SEQ ID NO: 208.

GGAAGCGGAGCTACTAACTTCAGCCTGCTGAAGCAGGCTGGAGACGTGGAGGAGAA
CCCTGGACCC (SEQ ID NO: 208)

**[0422]** In some embodiments, the HDRT polynucleotide comprises a coding sequence for a third (i.e., right or C-terminal) P2A peptide sequence as set forth in SEQ ID NO: 209.

GGCAGCGGCGCGACCAACTTTAGCCTGCTGAAACAGGCGGGCGATGTTGAAGAAAA
CCCGGGCCCG (SEQ ID NO: 209)

**[0423]** In some embodiments, the HDRT polynucleotide comprises a coding sequence for a 1928z-1XX CAR sequence as set forth in SEQ ID NO: 210.

ATGGCTCTCCCAGTGACTGCCCTACTGCTTCCCCTAGCGCTTCTCCTGCATGCAGAG
GTGAAGCTGCAGCAGTCTGGGGCTGAGCTGGTGAGGCCTGGGTCCTCAGTGAAGAT
TTCCTGCAAGGCTTCTGGCTATGCATTCAGTAGCTACTGGATGAACTGGGTGAAGCA
GAGGCCTGGACAGGGTCTTGAGTGGATTGGACAGATTTATCCTGGAGATGGTGATA
CTAACTACAATGGAAAGTTCAAGGGTCAAGCCACACTGACTGCAGACAAATCCTCC
AGCACAGCCTACATGCAGCTCAGCGGCCTAACATCTGAGGACTCTGCGGTCTATTTC
TGTGCAAGAAAGACCATTAGTTCGGTAGTAGATTTCTACTTTGACTACTGGGGCCAA
GGGACCACGGTCACCGTCTCCTCAGGTGGAGGTGGATCAGGTGGAGGTGGATCTGG
TGGAGGTGGATCTGACATTGAGCTCACCCAGTCTCCAAAATTCATGTCCACATCAGT
AGGAGACAGGGTCAGCGTCACCTGCAAGGCCAGTCAGAATGTGGGTACTAATGTAG
CCTGGTATCAACAGAAACCAGGACAATCTCCTAAACCACTGATTTACTCGGCAACCT
ACCGGAACAGTGGAGTCCCTGATCGCTTCACAGGCAGTGGATCTGGGACAGATTTC
ACTCTCACCATCACTAACGTGCAGTCTAAAGACTTGGCAGACTATTTCTGTCAACAA
TATAACAGGTATCCGTACACGTCCGGAGGGGGGACCAAGCTGGAGATCAAACGGGC
GGCCGCAATTGAAGTTATGTATCCTCCTCCTTACCTAGACAATGAGAAGAGCAATGG
AACCATTATCCATGTGAAAGGGAAACACCTTTGTCCAAGTCCCCTATTTCCCGGACC
TTCTAAGCCCTTTTGGGTGCTGGTGGTGGTTGGTGGAGTCCTGGCTTGCTATAGCTTG
CTAGTAACAGTGGCCTTTATTATTTTCTGGGTGAGGAGTAAGAGGAGCAGGCTCCTG
CACAGTGACTACATGAACATGACTCCCCGCCGCCCCGGGCCCACCCGCAAGCATTA
CCAGCCCTATGCCCCACCACGCGACTTCGCAGCCTATCGCTCCAGAGTGAAGTTCAG
CAGGAGCGCAGACGCCCCCGCGTACCAGCAGGGCCAGAACCAGCTCTATAACGAGC
TCAATCTAGGACGAAGAGAGGAGTACGATGTTTTGGACAAGAGACGTGGCCGGGAC
CCTGAGATGGGGGGAAAGCCGAGAAGGAAGAACCCTCAGGAAGGCCTGTTTAATGA
ACTGCAGAAAGATAAGATGGCGGAGGCCTTTAGTGAGATTGGGATGAAAGGCGAGC
GCCGGAGGGGCAAGGGGCACGATGGCCTTTTTCAGGGTCTCAGTACAGCCACCAAG
GACACCTTTGACGCCCTTCACATGCAGGCCCTGCCCCCTCGC (SEQ ID NO: 210)

**[0424]** In some embodiments, the HDRT polynucleotide comprises a coding sequence for an EGFRT polypeptide sequence as set forth in SEQ ID NO: 211.

ATGCTTCTCCTGGTGACAAGCCTTCTGCTCTGTGAGTTACCACACCCAGCATTCCTCC
TGATCCCACGCAAAGTGTGTAACGGAATAGGTATTGGTGAATTTAAAGACTCACTCT
CCATAAATGCTACGAATATTAAACACTTCAAAAACTGCACCTCCATCAGTGGCGATC
TCCACATCCTGCCGGTGGCATTTAGGGGTGACTCCTTCACACATACTCCTCCTCTGG
ACCCACAGGAACTGGATATTCTGAAAACCGTAAAGGAAATCACAGGGTTTTTGCTG
ATTCAGGCTTGGCCTGAAAACAGGACGGACCTCCATGCCTTTGAGAACCTAGAAATC
ATACGCGGCAGGACCAAGCAACATGGTCAGTTTTCTCTTGCAGTCGTCAGCCTGAAC
ATAACATCCTTGGGATTACGCTCCCTCAAGGAGATAAGTGATGGAGATGTGATAATT
TCAGGAAACAAAAATTTGTGCTATGCAAATACAATAAACTGGAAAAAACTGTTTGG
GACCTCCGGTCAGAAAACCAAAATTATAAGCAACAGAGGTGAAAACAGCTGCAAGG
CCACAGGCCAGGTCTGCCATGCCTTGTGCTCCCCCGAGGGCTGCTGGGGCCCGGAGC
CCAGGGACTGCGTCTCTTGCCGGAATGTCAGCCGAGGCAGGGAATGCGTGGACAAG
TGCAACCTTCTGGAGGGTGAGCCAAGGGAGTTTGTGGAGAACTCTGAGTGCATACA
GTGCCACCCAGAGTGCCTGCCTCAGGCCATGAACATCACCTGCACAGGACGGGGAC
CAGACAACTGTATCCAGTGTGCCCACTACATTGACGGCCCCCACTGCGTCAAGACCT
GCCCGGCAGGAGTCATGGGAGAAAACAACACCCTGGTCTGGAAGTACGCAGACGCC

GGCCATGTGTGCCACCTGTGCCATCCAAACTGCACCTACGGATGCACTGGGCCAGGT
CTTGAAGGCTGTCCCACGAATGGGCCTAAGATCCCGTCCATCGCCACTGGGATGGTG
GGGGCCCTCCTCTTGCTGCTGGTGGTGGCCCTGGGGATCGGCCTCTTCATG (SEQ ID NO: 211)

**[0425]** In some embodiments, the HDRT polynucleotide comprises the sequence as set forth in SEQ ID NO: 212 below, and e.g., as discussed in Examples 5, 7, and 11 below.

GATGTAAGGAGCTGCTGTGACTTGCTCAAGGCCTTATATCGAGTAAACGGTAGTGCT
GGGGCTTAGACGCAGGTGTTCTGATTTATAGTTCAAAACCTCTATCAATGAGAGAGC
AATCTCCTGGTAATGTGATAGATTTCCCAACTTAATGCCAACATACCATAAACCTCC
CATTCTGCTAATGCCCAGCCTAAGTTGGGGAGACCACTCCAGATTCCAAGATGTACA
GTTTGCTTTGCTGGGCCTTTTTCCCATGCCTGCCTTTACTCTGCCAGAGTTATATTGCT
GGGGTTTTGAAGAAGATCCTATTAAATAAAAGAATAAGCAGTATTATTAAGTAGCC
CTGCATTTCAGGTTTCCTTGAGTGGCAGGCCAGGCCTGGCCGTGAACGTTCACTGAA
ATCATGGCCTCTTGGCCAAGATTGATAGCTTGTGCCTGTCCCTGAGTCCCAGTCCATC
ACGAGCAGCTGGTTTCTAAGATGCTATTTCCCGTATAAAGCATGAGACCGTGACTTG
CCAGCCCCACAGAGCCCCGCCCTTGTCCATCACTGGCATCTGGACTCCAGCCTGGGT
TGGGGCAAAGAGGGAAATGAGATCATGTCCTAACCCTGGAATTGGATCCTCTTGTCT
TACAGATGGATCTGGAGCAACAAACTTCTCACTACTCAAACAAGCAGGTGACGTGG
AGGAGAATCCCGGCCCCATGGCTCTCCCAGTGACTGCCCTACTGCTTCCCCTAGCGC
TTCTCCTGCATGCAGAGGTGAAGCTGCAGCAGTCTGGGGCTGAGCTGGTGAGGCCTG
GGTCCTCAGTGAAGATTTCCTGCAAGGCTTCTGGCTATGCATTCAGTAGCTACTGGA
TGAACTGGGTGAAGCAGAGGCCTGGACAGGGTCTTGAGTGGATTGGACAGATTTAT
CCTGGAGATGGTGATACTAACTACAATGGAAAGTTCAAGGGTCAAGCCACACTGAC
TGCAGACAAATCCTCCAGCACAGCCTACATGCAGCTCAGCGGCCTAACATCTGAGG
ACTCTGCGGTCTATTTCTGTGCAAGAAAGACCATTAGTTCGGTAGTAGATTTCTACTT
TGACTACTGGGGCCAAGGGACCACGGTCACCGTCTCCTCAGGTGGAGGTGGATCAG
GTGGAGGTGGATCTGGTGGAGGTGGATCTGACATTGAGCTCACCCAGTCTCCAAAAT
TCATGTCCACATCAGTAGGAGACAGGGTCAGCGTCACCTGCAAGGCCAGTCAGAAT
GTGGGTACTAATGTAGCCTGGTATCAACAGAAACCAGGACAATCTCCTAAACCACT
GATTTACTCGGCAACCTACCGGAACAGTGGAGTCCCTGATCGCTTCACAGGCAGTGG
ATCTGGGACAGATTTCACTCTCACCATCACTAACGTGCAGTCTAAAGACTTGGCAGA
CTATTTCTGTCAACAATATAACAGGTATCCGTACACGTCCGGAGGGGGGACCAAGCT
GGAGATCAAACGGGCGGCCGCAATTGAAGTTATGTATCCTCCTCCTTACCTAGACAA
TGAGAAGAGCAATGGAACCATTATCCATGTGAAAGGGAAACACCTTTGTCCAAGTC
CCCTATTTCCCGGACCTTCTAAGCCCTTTTGGGTGCTGGTGGTGGTTGGTGGAGTCCT
GGCTTGCTATAGCTTGCTAGTAACAGTGGCCTTTATTATTTTCTGGGTGAGGAGTAA
GAGGAGCAGGCTCCTGCACAGTGACTACATGAACATGACTCCCCGCCGCCCCGGGC
CCACCCGCAAGCATTACCAGCCCTATGCCCCACCACGCGACTTCGCAGCCTATCGCT
CCAGAGTGAAGTTCAGCAGGAGCGCAGACGCCCCCGCGTACCAGCAGGGCCAGAAC
CAGCTCTATAACGAGCTCAATCTAGGACGAAGAGAGGAGTACGATGTTTTGGACAA
GAGACGTGGCCGGGACCCTGAGATGGGGGGAAAGCCGAGAAGGAAGAACCCTCAG
GAAGGCCTGTTTAATGAACTGCAGAAAGATAAGATGGCGGAGGCCTTTAGTGAGAT
TGGGATGAAAGGCGAGCGCCGGAGGGGCAAGGGGCACGATGGCCTTTTTCAGGGTC
TCAGTACAGCCACCAAGGACACCTTTGACGCCCTTCACATGCAGGCCCTGCCCCCTC

GCGGAAGCGGAGCTACTAACTTCAGCCTGCTGAAGCAGGCTGGAGACGTGGAGGAG
AACCCTGGACCCATGCTTCTCCTGGTGACAAGCCTTCTGCTCTGTGAGTTACCACAC
CCAGCATTCCTCCTGATCCCACGCAAAGTGTGTAACGGAATAGGTATTGGTGAATTT
AAAGACTCACTCTCCATAAATGCTACGAATATTAAACACTTCAAAAACTGCACCTCC
ATCAGTGGCGATCTCCACATCCTGCCGGTGGCATTTAGGGGTGACTCCTTCACACAT
ACTCCTCCTCTGGACCCACAGGAACTGGATATTCTGAAAACCGTAAAGGAAATCAC
AGGGTTTTTGCTGATTCAGGCTTGGCCTGAAAACAGGACGGACCTCCATGCCTTTGA
GAACCTAGAAATCATACGCGGCAGGACCAAGCAACATGGTCAGTTTTCTCTTGCAGT
CGTCAGCCTGAACATAACATCCTTGGGATTACGCTCCCTCAAGGAGATAAGTGATGG
AGATGTGATAATTTCAGGAAACAAAAATTTGTGCTATGCAAATACAATAAACTGGA
AAAAACTGTTTGGGACCTCCGGTCAGAAAACCAAAATTATAAGCAACAGAGGTGAA
AACAGCTGCAAGGCCACAGGCCAGGTCTGCCATGCCTTGTGCTCCCCCGAGGGCTG
CTGGGGCCCGGAGCCCAGGGACTGCGTCTCTTGCCGGAATGTCAGCCGAGGCAGGG
AATGCGTGGACAAGTGCAACCTTCTGGAGGGTGAGCCAAGGGAGTTTGTGGAGAAC
TCTGAGTGCATACAGTGCCACCCAGAGTGCCTGCCTCAGGCCATGAACATCACCTGC
ACAGGACGGGGACCAGACAACTGTATCCAGTGTGCCCACTACATTGACGGCCCCCA
CTGCGTCAAGACCTGCCCGGCAGGAGTCATGGGAGAAAACAACACCCTGGTCTGGA
AGTACGCAGACGCCGGCCATGTGTGCCACCTGTGCCATCCAAACTGCACCTACGGAT
GCACTGGGCCAGGTCTTGAAGGCTGTCCCACGAATGGGCCTAAGATCCCGTCCATCG
CCACTGGGATGGTGGGGGCCCTCCTCTTGCTGCTGGTGGTGGCCCTGGGGATCGGCC
TCTTCATGGGCAGCGGCGCGACCAACTTTAGCCTGCTGAAACAGGCGGGCGATGTTG
AAGAAAACCCGGGCCCGAATATCCAGAACCCTGACCCTGCCGTGTACCAGCTGAGA
GACTCTAAATCCAGTGACAAGTCTGTCTGCCTATTCACCGATTTTGATTCTCAAACA
AATGTGTCACAAAGTAAGGATTCTGATGTGTATATCACAGACAAAACTGTGCTAGAC
ATGAGGTCTATGGACTTCAAGAGCAACAGTGCTGTGGCCTGGAGCAACAAATCTGA
CTTTGCATGTGCAAACGCCTTCAACAACAGCATTATTCCAGAAGACACCTTCTTCCC
CAGCCCAGGTAAGGGCAGCTTTGGTGCCTTCGCAGGCTGTTTCCTTGCTTCAGGAAT
GGCCAGGTTCTGCCCAGAGCTCTGGTCAATGATGTCTAAAACTCCTCTGATTGGTGG
TCTCGGCCTTATCCATTGCCACCAAAACCCTCTTTTTACTAAGAAACAGTGAGCCTTG
TTCTGGCAGTCCAGAGAATGACACGGGAAAAAAGCAGATGAAGAGAAGGTGGCAG
GAGAGGGCACGTGGCCCAGCCTCAGTCTCTCCAACTGAGTTCCTGCCTGCCTGCCTT
TGCTCAGACTGTTTGCCCCTTACTGCTCTTCTAGGCCTCATTCTAAGCCCCTTCTCCA
AGTTG (SEQ ID NO: 212)

## V. Polypeptide Linkers

**[0426]** In some embodiments, a viral envelope protein, a targeting polypeptide, a CRISPR-Cas polypeptide, or a cell receptor can be fused to polypeptide, e.g., a fusion partner, via a linker polypeptide (e.g., one or more linker polypeptides). The linker polypeptide may have any of a variety of amino acid sequences. Proteins can be joined by a spacer peptide, generally of a flexible nature, although other chemical linkages are not excluded. Suitable linkers include polypeptides of between 4 amino acids and 40 amino acids in length, or between 4 amino acids and 25 amino acids in length. These linkers can be produced by using synthetic, linker-encoding oligonucleotides

to couple the proteins, or can be encoded by a nucleic acid sequence encoding the fusion protein. Peptide linkers with a degree of flexibility can be used. The linking peptides may have virtually any amino acid sequence, bearing in mind that the preferred linkers will have a sequence that results in a generally flexible peptide. The use of small amino acids, such as glycine and alanine, are of use in creating a flexible peptide. The creation of such sequences is routine to those of skill in the art. A variety of different linkers are commercially available and are considered suitable for use.

**[0427]** Examples of linker polypeptides include glycine polymers (G)n where n is an integer of at least one; glycine-serine polymers (including, for example, (GS)n, (GSGGS)n (SEQ ID NO: 213), (GGSGGS)n (SEQ ID NO: 214), (GGGGS)n (SEQ ID NO: 215), and (GGGS)n (SEQ ID NO: 216), where n is an integer of at least one; e.g., where n is an integer from 1 to 10); glycine-alanine polymers; and alanine-serine polymers. Exemplary linkers can comprise amino acid sequences including, but not limited to, GGSG (SEQ ID NO: 217), GGSGG (SEQ ID NO: 218), GSGSG (SEQ ID NO: 219), GSGGG (SEQ ID NO: 220), GGGSG (SEQ ID NO: 221), GSSSG (SEQ ID NO: 222), GGGGS (SEQ ID NO: 223), and the like. The ordinarily skilled artisan will recognize that design of a peptide conjugated to any desired element can include linkers that are all or partially flexible, such that the linker can include a flexible linker as well as one or more portions that confer less flexible structure.

## VI. Methods for Modifying Cells

**[0428]** Also provided herein are for methods of introducing at least two vectors of the present disclosure, e.g., an AAV vector with an EDV or an LNP, to a subject, e.g., a mammal, for modifying a gene in the subject. In some embodiments, the method comprises *in vivo* DNA insertion into a gene in a target cell of the subject. In some embodiments, the method comprises administering to the subject a composition comprising the vectors. In some embodiments, methods herein include delivering the vectors to a target cell, comprising contacting the target cell with the vectors. In some embodiments, the one or more vectors can comprise a viral envelope protein (e.g., a VSVG or a variant thereof), a targeting polypeptide, a polynucleotide-guided nuclease (e.g., a CRISPR-Cas nuclease), a guide polynucleotide (e.g., a gRNA), a virus capsid (e.g., an AAV capsid or a variant thereof), a donor template polynucleotide, or any combination thereof. In some embodiments, the one or more vectors delivers the polynucleotide-guided nuclease (e.g., a

CRISPR-Cas nuclease), the guide polynucleotide (e.g., a gRNA), a virus capsid (e.g., an AAV capsid or a variant thereof), and/or the donor template polynucleotide to the nucleus of a target cell, e.g., T cell, B cell, NK cell, monocyte, macrophage, dendritic cell, or HSC nucleus.

**[0429]** Because the above-mentioned gene-editing components are delivered by at least two vectors to a target cell, *in vivo* gene modification takes place only when all the relevant vectors are used and all the gene-editing components are present together in a target cell. For example, a polynucleotide-guided nuclease (e.g., a CRISPR-Cas nuclease), a guide polynucleotide (e.g., a gRNA), and a donor template polynucleotide must be present in the same target cell nucleus at the same time to produce cleavage at a target gene and integration of the donor template polynucleotide into the target gene. In some embodiments, the polynucleotide-guided nuclease (e.g., a CRISPR-Cas nuclease) and the guide polynucleotide (e.g., a gRNA) are delivered by an EDV or an LNP, while the donor template polynucleotide is delivered by an AAV vector. In some embodiments, the polynucleotide-guided nuclease (e.g., a CRISPR-Cas nuclease) is delivered by an EDV or an LNP, while the guide polynucleotide (e.g., a gRNA) and the donor template polynucleotide are delivered by an AAV vector. In some embodiments, the polynucleotide-guided nuclease (e.g., a CRISPR-Cas nuclease) and the guide polynucleotide (e.g., a gRNA) are delivered by an EDV or an LNP, while the guide polynucleotide (e.g., a gRNA) and the donor template polynucleotide is delivered by an AAV vector. Because both EDV/LNP and AAV vectors must be present, and because each vector has a different mechanism for cell target selectivity, the methods herein for genome modification are highly specific.

**[0430]** In some embodiments, the method comprises contacting the vectors with a target cell, e.g., T cell (CD4 T cell and/or CD8 T cell), a B cell, NK cell, mast cell, dendritic cell, macrophage, monocyte, HSC, or any combination thereof. In some embodiments, the vectors can be delivered to a specific tissue by administering one or more vectors with a targeting polypeptide and/or with one or more vectors with enhanced tropism to a T cell (CD4 T cell and/or CD8 T cell), a B cell, a NK cell, a mast cell, a dendritic cell, a macrophage, monocyte, HSC, or any combination thereof.

**[0431]** In some embodiments, the methods of administering the vectors modify expression of at least one native protein and/or native gene in a target cell as compared to baseline. As used herein, "baseline" refers to the expression of the native protein and/or native gene before the one or more vectors were administered. As used herein, "substantially modifies expression" refers to at least a

1-fold change in expression (e.g., decreased expression) as compared to baseline. In some embodiments, methods of administering the vectors modify expression of the protein and/or gene as compared to baseline by at least about 2-fold to about 50-fold (e.g., about 2-, 4-, 6-, 8-, 10-, 20-, 30-, 40-, 50-fold). In some embodiments, the methods of administering the vectors to a subject modify expression of the one protein and/or gene as compared to baseline by about 1% to about 100%, about 5% to about 95%, about 10% to about 90%, about 15% to about 85%, or about 20% to about 80%.

## I. Methods of Treatment

**[0432]** In some embodiments, an effective amount of the vectors disclosed herein can be given to a subject in need thereof to alleviate one or more symptoms associated with a disease and or condition. In some embodiments, the effective amount confers a therapeutic effect on a subject having a disease and or condition. In some embodiments, an effective amount can be an amount that reduces at least one symptom of disease or condition in the subject.

### a. Diseases

**[0433]** Certain antigens overexpressed in cells implicated in disease but are minimally expression in normal tissue. In some embodiments, the vectors of the present disclosure may be used to treat patients with a disease that is associated with a detectable disease-related antigen, including a cancer antigen or a tumor antigen.

**[0434]** In some embodiments, a patient in need thereof can have been diagnosed with a cancer. As used herein, the term "cancer" encompasses precancerous, neoplastic, transformed, and cancerous cells, and can refer to a solid tumor, or a non-solid cancer. Cancer includes both benign and malignant neoplasms (abnormal growth). The term "cancer" can thus refer to carcinomas, sarcomas, adenocarcinomas, lymphomas, leukemias, solid and lymphoid cancers, etc. Exemplary cancer antigens and tumor antigens are discussed in detail above. Non-limiting examples of cancer include breast cancers; sarcomas (including but not limited to osteosarcomas and angiosarcomas and fibrosarcomas); leukemias and lymphomas (including acute myelogenous leukemia (AML), B-cell acute lymphoblastic leukemia (B-ALL), chronic lymphocytic leukemia (CLL or B-CLL), hairy cell leukemia, B-cell prolymphocytic leukemia, Non-Hodgkin lymphoma or Non-Hodgkin's lymphoma, Hodgkin lymphoma or Hodgkin's lymphoma, multiple myeloma (MM), or T-cell

leukemia, diffuse large B-cell lymphoma (DLBCL), follicular lymphoma, follicle center lymphoma, mantle cell lymphoma, Burkitt lymphoma (BL), Waldenstrom macroglobulinemia, anaplastic large cell lymphoma, peripheral T cell lymphoma, cutaneous T cell lymphoma, extranodal natural killer (NK)/T-cell lymphoma, Epstein-Barr virus associated T cell lymphoma, or T-cell acute lymphoblastic leukemia); genitourinary cancers (including but not limited to ovarian, urethral, bladder (e.g., bladder urothelial carcinoma), and prostate cancers); gastrointestinal cancers (including but not limited to colon esophageal and stomach cancers); lung cancers (non-small cell lung cancer (NSCLC), small cell lung cancer); myelomas; thyroid cancer; pleural cancer; pancreatic cancers; liver cancers; kidney cancers; HPV-negative head and neck squamous cell carcinoma (HNSCC); a solid malignancy that is microsatellite high (MSI H)/mismatch repair (MMR) deficient; renal cancer; gastric cancer; tumor mutational burden high tumors; endocrine cancers; skin cancers; and brain or central and peripheral nervous (CNS) system tumors, malignant or benign cancers, including gliomas and neuroblastomas, astrocytomas, myelodysplastic disorders; cervical carcinoma-in-situ; intestinal polyposes; oral leukoplakias; and histiocytoses.

**[0435]** In some embodiments, a patient in need thereof can have been diagnosed with an infectious disease. By example, but not limited to, a patient can have been diagnosed with chickenpox, common cold, diphtheria, *E. coli*, giardiasis, HIV/AIDS, infectious mononucleosis, influenza, Lyme disease, malaria, measles, meningitis, mumps, poliomyelitis (polio), pneumonia, Rocky mountain spotted fever, rubella (German measles), Salmonella infections, severe acute respiratory syndrome (SARS), sexually transmitted diseases, shingles (herpes zoster), tetanus, toxic shock syndrome, tuberculosis, viral hepatitis, West Nile virus, whooping cough (pertussis), or a combination thereof.

**[0436]** In some embodiments, a patient in need thereof can have been diagnosed with an autoimmune disease. By example, but not limited to, a patient can have been diagnosed with diabetes (Type 1), lupus, multiple sclerosis, rheumatoid arthritis, celiac disease, or a combination thereof.

**[0437]** In some embodiments, a patient in need thereof can have been diagnosed with an immune deficiency disease. By example, but not limited to, a patient can have been diagnosed with autoimmune lymphoproliferative syndrome (ALPS), autoimmune polyglandular syndrome type 1

(APS-1), BENTA disease, caspase eight deficiency state (CEDS), CARD9 deficiency and other syndromes of susceptibility to *Candidiasis*, chronic granulomatous disease (CGD), common variable immunodeficiency (CVID), congenital neutropenia syndromes, CTLA4 deficiency, DOCK8 deficiency, *GATA2* deficiency, hyper-immunoglobulin E syndrome (HIES), hyper-immunoglobulin M (IgM) syndrome, leukocyte adhesion deficiency (LAD), *LRBA* deficiency, PI3 kinase disease, PLAID and/or PLAID-like disease, severe combined immunodeficiency (SCID), *STAT3* gain-of-function disease, Warts, Hypogammaglobulinemia, Infections, and Myelokathexis (WHIM) Syndrome, Wiskott-Aldrich syndrome (WAS), X-linked agammaglobulinemia (XLA), XMEN disease, lupus or Type I or Type II diabetes or a combination thereof.

**b. Doses**

**[0438]** In some embodiments, a therapeutically effective amount of the disclosed vectors can comprise a range of about $1x10^8$ vg/kg to about $1x10^{16}$ vg/kg. In some embodiments, an AAV vector is administered at a dose of about $1x10^8$ to about $1x10^{14}$ vg/kg, about $1x10^{11}$ to about $1x10^{16}$ vg/kg, or about $1x10^{11}$ to about $1x10^{14}$ vg/kg. In some embodiments, the AAV vector is administered at $1x10^{11}$ to $1x10^{14}$ vg/kg. In some embodiments, an EDV vector is administered at a dose of about $1x10^8$ to about $1x10^{13}$ vg/kg, about $1x10^{10}$ to about $1x10^{16}$ vg/kg, or about $1x10^{10}$ to about $1x10^{13}$ vg/kg. In some embodiments, the EDV vector is administered at $1x10^{10}$ to $1x10^{13}$ vg/kg. In some embodiments, for example, an LNP vector is administered at a dose of about $1x10^8$ to about $1x10^{16}$ vg/kg, about $1x10^{10}$ to about $1x10^{16}$ vg/kg, or about $1x10^{10}$ to about $1x10^{13}$ vg/kg. In some embodiments, the LNP vector is administered at $1x10^{10}$ to $1x10^{13}$ vg/kg.

**[0439]** In some embodiments, the methods comprise administering the vectors to a subject at least once. In some embodiments, the methods include administering the vectors to a subject more than once. In some embodiments, the methods include administering the vectors to a subject between at least once to at least 10 times (e.g., about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 times). In some embodiments, the methods include administering the vectors at least twice, at least 3 times, at least 4 times, or at least 5 times. In some embodiments, the methods include administering the vectors to a subject once a day, once every other day, once a week, once every two weeks, once every three weeks, once a month, once every other month, once every three months, once every four months, once a year, or twice a year. In some embodiments, the methods include administering the vector(s) at many times as needed to see the desired response. In some embodiments, the

desired response can be attenuation of at least one symptom of a disease and/or condition in a subject after administration of a dose of the vectors compared to before administration of the vector(s).

**[0440]** In some embodiments, the methods of administering the vectors produces expression of a heterologous polypeptide, e.g., a CAR. In these embodiments, the heterologous polypeptide is a polypeptide that is not native to the target cell.

## II. Compositions

**[0441]** In some embodiments, any of the disclosed vectors (e.g., AAV vectors, EDVs, and LNPs) can be formulated to form a pharmaceutical composition. In some embodiments, pharmaceutical compositions herein can further include a pharmaceutically acceptable carrier, diluent, or excipient. Any of the pharmaceutical compositions to be used in the presently disclosed methods can comprise pharmaceutically acceptable carriers, excipients, or stabilizers in the form of lyophilized formations or aqueous solutions.

**[0442]** The carrier in the pharmaceutical composition must be "acceptable" in the sense that it is compatible with the active ingredient of the composition, and preferably, capable of stabilizing the active ingredient and not deleterious to the subject to be treated. For example, "pharmaceutically acceptable" can refer to molecular entities and other ingredients of compositions comprising such that are physiologically tolerable and do not typically produce untoward reactions when administered to a mammal (e.g., a mouse, a primate, or a human). In some embodiments, the "pharmaceutically acceptable" carrier used in the pharmaceutical compositions disclosed herein can be those approved by a regulatory agency of the Federal or a state government or listed in the U.S. Pharmacopeia or other generally recognized pharmacopeia for use in mammals, and more particularly in humans.

**[0443]** Pharmaceutically acceptable carriers, including buffers, are well known in the art, and can comprise phosphate, citrate, and other organic acids; antioxidants including ascorbic acid and methionine; preservatives; low molecular weight polypeptides; proteins, such as serum albumin, gelatin, or immunoglobulins; amino acids; hydrophobic polymers; monosaccharides; disaccharides; and other carbohydrates; metal complexes; and/or non-ionic surfactants. See, e.g.

Remington: The Science and Practice of Pharmacy 20th Ed. (2000) Lippincott Williams and Wilkins, Ed. K. E. Hoover.

**[0444]** In some embodiments, the pharmaceutical compositions or formulations herein are for parenteral administration, such as intravenous, intracerebroventricular injection, intra-cisterna magna injection, intra-parenchymal injection, or a combination thereof. Such pharmaceutically acceptable carriers can be sterile liquids, such as water and oil, including those of petroleum, animal, vegetable or synthetic origin, such as peanut oil, soybean oil, mineral oil, and the like. Saline solutions and aqueous dextrose, polyethylene glycol (PEG) and glycerol solutions can also be employed as liquid carriers, particularly for injectable solutions. Pharmaceutical compositions disclosed herein can further comprise additional ingredients, for example preservatives, buffers, tonicity agents, antioxidants and stabilizers, nonionic wetting or clarifying agents, viscosity-increasing agents, and the like. The pharmaceutical compositions described herein can be packaged in single unit dosages or in multidosage forms.

**[0445]** Formulations suitable for parenteral administration include aqueous and non-aqueous sterile injection solutions which can contain anti-oxidants, buffers, bacteriostats and solutes which render the formulation isotonic with the blood of the intended recipient; and aqueous and non-aqueous sterile suspensions which can include suspending agents and thickening agents. Aqueous solutions can be suitably buffered (preferably to a pH of from 3 to 9). The preparation of suitable parenteral formulations under sterile conditions is readily accomplished by standard pharmaceutical techniques well known to those skilled in the art.

**[0446]** The pharmaceutical compositions to be used for *in vivo* administration should be sterile. This is readily accomplished by, for example, filtration through sterile filtration membranes. Sterile injectable solutions are generally prepared by incorporating the active components (e.g., EDVs, AAV vectors, virus capsids, AAV viral particles, and LNPs in the required amounts in the appropriate solvent with various other ingredients enumerated above, as required, followed by filter sterilization. Generally, dispersions are prepared by incorporating the sterilized active ingredient into a sterile vehicle which contains the basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, the preferred methods of preparation are vacuum drying and the freeze-

drying technique that yield a powder of the active ingredient plus any additional desired ingredient from the previously sterile-filtered solution thereof.

**[0447]** The pharmaceutical compositions disclosed herein can also comprise other ingredients such as diluents and adjuvants. Acceptable carriers, diluents and adjuvants are nontoxic to recipients and are preferably inert at the dosages and concentrations employed, and include buffers such as phosphate, citrate, or other organic acids; antioxidants such as ascorbic acid; low molecular weight polypeptides; proteins, such as serum albumin, gelatin, or immunoglobulins; hydrophilic polymers such as polyvinylpyrrolidone; amino acids such as glycine, glutamine, asparagine, arginine or lysine; monosaccharides, disaccharides, and other carbohydrates including glucose, mannose, or dextrins; chelating agents such as EDTA; sugar alcohols such as mannitol or sorbitol; salt-forming counterions such as sodium; and/or nonionic surfactants such as Tween, pluronics, or polyethylene glycols.

**a. Exemplary Compositions**

**[0448]** In some embodiments, a composition comprises: (a) a first vector that delivers a polynucleotide-guided nuclease to T cell nuclei; (b) a second vector comprising a donor template polynucleotide; (c) a guide polynucleotide for the polynucleotide-guided nuclease; and (d) a pharmaceutical carrier and/or a pharmaceutical excipient. In some embodiments, the guide polynucleotide is targeted to a portion of a gene in the T cell genome that is selectively expressed in T cells.

**[0449]** In some embodiments, the guide polynucleotide for the polynucleotide-guided nuclease is present in the first vector. In some embodiments, the guide polynucleotide for the polynucleotide-guided nuclease is present in the second vector. In some embodiments, the guide polynucleotide for the polynucleotide-guided nuclease is present in both the first and second vectors.

**[0450]** In some embodiments, the first vector is an EDV. In some embodiments, the first vector is an LNP. In some embodiments, the second vector is an AAV vector. EDVs, LNPs, and AAV vectors are discussed above in detail. In some embodiments, the EDV or LNP comprises the VSVG viral envelope protein with the sequence as set forth in SEQ ID NO: 81. In some embodiments, the EDV or LNP comprises the viral envelope protein VSVGmut with the sequence as set forth in

SEQ ID NO: 81. In some embodiments, the AAV vector comprises a capsid variant with the sequence HAPRVEE (SEQ ID NO: 3) at positions corresponding to amino acids 454-460 (VP1 numbering) of a native AAV6 capsid protein, (SEQ ID NO: 1).

**[0451]** In some embodiments, the polynucleotide-guided nuclease is Cas9. In some embodiments, the guide polynucleotide is a gRNA. Cas9 and other suitable polynucleotide-guided nucleases and gRNAs are discussed above in detail.

**[0452]** In some embodiments, the composition comprises (a) an EDV vector that comprises delivers a Cas9 to T cell nuclei; (b) an AAV vector comprising a donor template polynucleotide; (c) a gRNA for the Cas9; and (d) a pharmaceutical carrier and/or a pharmaceutical excipient. In some embodiments, the EDV, the AAV vector, or both the EDV and the AAV vector comprise the gRNA.

**[0453]** In some embodiments, the composition comprises (a) an LNP vector that comprises delivers a Cas9 to T cell nuclei; (b) an AAV vector comprising a donor template polynucleotide; (c) a gRNA for the Cas9; and (d) a pharmaceutical carrier and/or a pharmaceutical excipient. In some embodiments, the LNP, the AAV vector, or both the LNP and the AAV vector comprise the gRNA.

**[0454]** In some embodiments, the gRNA comprises the sequence of CAGGGTTCTGGATATCTGT (SEQ ID NO: 137) or TCAGGGTTCTGGATATCTGT (SEQ ID NO: 138). In some embodiments, the gRNA is under control of the U6 promoter. In some embodiments, the EDV, LNP, and/or AAV vector comprises a U6 promoter-TRAC-sgRNA scaffold comprises the sequence as set forth in SEQ ID NO: 139.

**[0455]** In some embodiments, the donor template polynucleotide of the AAV vector comprises a coding sequence for a polypeptide comprising (a) an extracellular target-binding domain, (b) a transmembrane domain, (c) a hinge domain, and (d) an intracellular signaling domain. donor template polynucleotides and heterologous polypeptides are discussed above in detail. In some embodiments, (a) the polypeptide extracellular target-binding domain binds to CD19 polypeptide, (b) the polypeptide transmembrane domain and/or hinge domain are derived form a CD28 polypeptide, and (c) the polypeptide intracellular signaling domain comprises a modified CD3ζ polypeptide comprising (a) a native ITAM1 (SEQ ID NO: 153), (b) a modified ITAM2 comprising

two Tyr to Phe mutations (SEQ ID NO: 156), (c) a modified ITAM3 comprising two Tyr to Phe mutations (SEQ ID NO: 158), (d) a native BRS1, (e) a native BRS2, (f) a native BRS3, and (g) a co-stimulatory signaling region comprising a CD28 polypeptide. In some embodiments, the polypeptide is CAR designated as "1928z-1XX" (SEQ ID NO: 160) and the donor template polynucleotide comprises a coding sequence for the 1928z-1XX CAR (SEQ ID NO: 210). In some embodiments, the donor template polynucleotide also comprises a coding sequence for EGFRT polypeptide (SEQ ID NO: 211). In some embodiments, the donor template polynucleotide comprises the sequence as set forth in SEQ ID NO: 212.

**[0456]** In some embodiments, (1) the first vector is an EDV as described herein, for example, comprising (a) the viral envelope protein (e.g., VSVGmut (SEQ ID NO: 81)) and (b) the polynucleotide-guided nuclease CRISPR-Cas9; and (2) the second vector is an AAV as described herein, for example, comprising (a) an AAV6 capsid or variant thereof (e.g., with the peptide sequence HAPRVEE (SEQ ID NO: 3) at positions corresponding to amino acids 454-460 (VP1 numbering) of a native AAV6 capsid protein (SEQ ID NO: 1)) and (b) an HDRT polynucleotide comprising (i) a first TRAC locus homology sequence (e.g., SEQ ID NO: 205), (ii) a second TRAC locus homology sequence (e.g., SEQ ID NO: 206), and (iii) a coding sequence for a CAR designated as "19-166-28z 1XX" or "1928z-1XX" (SEQ ID NO: 210). In some of these embodiments, the EDV or the AAV further comprises (a) a gRNA comprising the sequence of SEQ ID NO: 137, (b) a gRNA comprising the sequence of SEQ ID NO: 138, or (c) a U6 promoter-TRAC-sgRNA scaffold (SEQ ID NO: 139), or is otherwise a gRNA directed to the TRAC locus, e.g., exon 1 of the TRAC locus.

**[0457]** In some embodiments, (1) the first vector is an EDV as described herein, for example, comprising (a) the viral envelope protein (e.g., VSVGmut (SEQ ID NO: 81)) and (b) the polynucleotide-guided nuclease CRISPR-Cas9; and (2) the second vector is an AAV as described herein, for example, comprising (a) an AAV6 capsid or variant thereof (e.g., with the peptide sequence HAPRVEE (SEQ ID NO: 3) at positions corresponding to amino acids 454-460 (VP1 numbering) of a native AAV6 capsid protein (SEQ ID NO: 1)) and (b) an HDRT polynucleotide comprising (i) a first TRAC locus homology sequence (e.g., SEQ ID NO: 205), (ii) a second TRAC locus homology sequence (e.g., SEQ ID NO: 206), and (iii) a coding sequence for a CAR designated as "19-166-28z 1XX" or "1928z-1XX" (SEQ ID NO: 210), and (iv) a coding sequence

for an EGFRT polypeptide (SEQ ID NO: 211). In some of these embodiments, the EDV or the AAV further comprises (a) a gRNA comprising the sequence of SEQ ID NO: 137, (b) a gRNA comprising the sequence of SEQ ID NO: 138, or (c) a U6 promoter-TRAC-sgRNA scaffold (SEQ ID NO: 139), or is otherwise a gRNA directed to the TRAC locus, e.g., exon 1 of the TRAC locus.

**[0458]** In some embodiments, (1) the first vector is an EDV as described herein, for example, comprising (a) the viral envelope protein (e.g., VSVGmut (SEQ ID NO: 81)) and (b) the polynucleotide-guided nuclease CRISPR-Cas9; and (2) the second vector is an AAV as described herein, for example, comprising (a) an AAV6 capsid or variant thereof (e.g., with the peptide sequence HAPRVEE (SEQ ID NO: 3) at positions corresponding to amino acids 454-460 (VP1 numbering) of a native AAV6 capsid protein (SEQ ID NO: 1)) and (b) an HDRT polynucleotide comprising (i) a first TRAC locus homology sequence (e.g., SEQ ID NO: 205), (ii) a second TRAC locus homology sequence (e.g., SEQ ID NO: 206), and (iii) coding sequences for a CAR and an EGFRT designated as "1928z-1XX-P2A-EGFRT" (SEQ ID NO: 212). In some of these embodiments, the EDV or the AAV further comprises (a) a gRNA comprising the sequence of SEQ ID NO: 137, (b) a gRNA comprising the sequence of SEQ ID NO: 138, or (c) a U6 promoter-TRAC-sgRNA scaffold (SEQ ID NO: 139), or is otherwise a gRNA directed to the TRAC locus, e.g., exon 1 of the TRAC locus.

## EXAMPLES

**[0459]** The following examples are provided by way of illustration only and not by way of limitation. Those of skill in the art will readily recognize a variety of non-critical parameters that could be changed or modified to yield essentially the same or similar results.

### Overview for Examples 1-10 – *In Vivo* Generation of TRAC-CAR T Cells Using Enveloped Delivery Vehicles (EDVs)

**[0460]** The Examples 1-10 below relate to *in vivo* methods using specific AAVs and CD3 targeting enveloped delivery vehicles (EDVs) loaded with Cas9 nuclease to precisely integrate CAR transgenes to the *TRAC* locus of T cells. *In vivo* generated *TRAC*-CAR T cells would combine T cell specific and physiological CAR expression while bypassing the *ex vivo* cell manufacturing and patients pre-conditioning.

**Example 1 – Materials and Methods**

**[0461]** The materials and methods for Examples 2-7 below are provided here in Example 1.

**[0462]** **Enveloped delivery vehicles (EDVs).** The following EDVs were used in the Examples herein to deliver Cas9/sg*TRAC* ribonucleoprotein (RNP) to a cell. A first generation EDV was used that was coated with a WT vesicular stomatitis virus glycoprotein G (WT-VSVG EDV). WT-VSVG EDV is discussed in Examples 2-3 and Figures 1A-1D. A second generation EDV was used that was coated with a mutated VSVG (mVSVG contains K47Q and R354A; SEQ ID NO: 81) that has fusogenic activity and an anti-CD3 antibody (VSVGm-aCD3 EDV). VSVGm-aCD3 EDV is discussed in Examples 4-7 and Figures 2A-2D and 3A-3D.

**[0463]** **AAV capsids.** The following AAV capsids were used in the Examples herein to deliver a homology-directed repair template (HDRT). AAV6 is an AAV capsid with a WT serotype known for its broad tropism, including T cells, natural killer (NK) cells, B cells and HSCs. AAV-Ark312 (Ark312) is an AAV capsid that was evolved *ex vivo* on primary T cells to resist neutralizing antibodies and to preferably infect T cells. AAV6 is discussed in Examples 2-7 and Ark312 is discussed in Examples 3-4.

**[0464]** **CD-19-Targeting *TRAC*-CAR T Cells.** To generate CD19-targeting *TRAC*-CAR T cells, a nucleic acid sequence encoding a 1928z-1XX CAR (SEQ ID NO: 210) was used in certain HDRTs discussed below. The 1928z-1XX CAR (SEQ ID NO: 160) was previously described in the U.S. Patent Application Publication No. 2020/0317777.

**[0465]** ***Ex vivo* transduction of human T cells.** In general, human T cells were activated for 48 hours using anti-CD3/CD28 dynabeads. The cells were then transduced with various combinations of concentrated EDVs carrying Cas9/sg*TRAC* RNPs with AAVs carrying an HDRTs.

**[0466]** **Flow cytometry.** Cell expression of TCRs, EGFR, and cell surface markers (e.g., CD4, CD8, CD19, CD25, and CD45) were determined by flow cytometry. Flow cytometry analysis was typically performed 72 to 96 hours after T cell transduction.

**[0467]** **Cytotoxicity assay.** Cell cytotoxicity was measured using a peripheral blood mononuclear cell (PBMC) cytotoxicity luciferase assay kit for toxicity towards a NALM6 cell line. NALM6 is a CD19+ acute lymphoblastic leukemia (ALL) cell line. In general, the method

produces a cytotoxic activity profile of target PBMCs, e.g., transduced T cells, towards the firefly luciferase NALM6 cell line – a decrease in luciferase signal indicates toxicity of the target PBMCs towards NALM6. NALM6 cells were co-cultured with T cells for 24 hours with several effector to tumor cell (E:T) ratios.

**[0468]** **A humanized mouse model.** Six to 8-week old immunodeficient NOD/SCID/IL2Rγ-/- (NSG) mice were acquired from the Jackson Laboratory (jax.org). The mice were engrafted with human PBMCs for reconstitution of T cell, NK cell, and B cell populations in the mice (Figure 3A). Two weeks after human PBMC engraftment, the mice receive intravenous (IV) injections of mixtures of $5x10^{10}$ particles of (1) WT-VSVG EDV or VSVGm-aCD3 EDV carrying Cas9/sg*TRAC* and (2) $1x10^{12}$ particles of AAV or Ark312 carrying a 1928z-1XX-P2A-EGFRT HDRT (SEQ ID NO: 212). Control mice were injected with phosphate buffer saline (PBS). Two weeks later, the mice were euthanized and the organs, including the spleen, were harvested. Mice spleens were analyzed by flow cytometry: B cells were determined as CD45+/CD19+ and CAR-T cells were determined as CD45+/CAR+/EGFRT+.

### Example 2 – Transduction of T Cells with a WT-VSVG EDV Carrying a *TRAC*-Targeting Cas9 RNP

**[0469]** This Example demonstrates the use of the first generation EDVs to deliver a *TRAC*-targeting Cas9 RNP (Cas9/sg*TRAC* RNP) to T cells. The first generation EDV was coated with WT-VSVG (WT-VSVG EDV; labeled "Cas9-sg*TRAC*" in Figure 1A). Human T cells were transduced *in vitro* with 50 µl of EDVs at different MOIs: (1) $3\times10^5$ T cells (low MOI), (2) $2\times10^5$ T cells (medium MOI), and (3) $1\times10^5$ T cells (high MOI). The cells were transduced with EDVs in combination with WT AAV6 carrying an HDRT (labeled "*TRAC*-CAR-sg*TRAC*" in Figure 1A). Seventy-two hours after transduction, the cells were analyzed by flow cytometry.

**[0470]** As shown in Figure 1B, the first-generation WT-VSVG EDV successfully delivered a Cas9/sg*TRAC* RNP in activated primary human T cells and disrupted TCR expression with the AAV6-delivered HDRT. TCR expression was knocked-down in a dose-response manner; a 60% decrease of TCR expression was achieved.

**Example 3 - Transduction of T Cells with WT-VSVG EDV and AAV-Ark312**

**[0471]** This Example demonstrates the use of an AAV with an Ark312 variant (SEQ ID NO: 58) to deliver a *TRAC*-CAR HDRT to T cells. Ark312 was evolved *ex vivo* for resistance against neutralization by antibodies and for the ability to preferably infect T cells. $1\times10^5$ activated human T cells were transduced *in vitro* with either (1) Ark312 or AAV6 in combination with (2) 50 µl WT-VSVG EDV carrying Cas9/sg*TRAC* RNPs. Seventy-two hours after transduction, the cells were analyzed by flow cytometry.

**[0472]** As shown in Figures 1C and 1D, both AAV6 and Ark312 serotypes show EGFR CAR expression, indicative of *TRAC* CAR knock-in. The EDVs were administered at three different doses at 50 µl each: (1) $3\times10^5$ T cells (low MOI), (2) $2\times10^5$ T cells (medium MOI), and (3) $1\times10^5$ T cells (high MOI). As shown in Figure 1D, average knock-in efficiency of EGFRT expression was 43% with AAV6 and 25% with Ark312. These efficiency numbers were comparable to efficiencies previously reported for *TRAC*-CAR T cells (Eyquem, J. et al. Targeting a CAR to the *TRAC* locus with CRISPR-Cas9 enhances tumor rejection. Nature 543, 113–117 (2017)).

**Example 4 - Transduction of T Cells using an EDV with a Fusogenic Variant and an Anti-CD3 Antibody**

**[0473]** This Example demonstrates the use of a second generation EDV which comprises two features on its surface: (1) a VSVG variant with fusogenic activity (VSVGm), and (2) a CD3-targeting antibody (aCD3). In contrast with the WT-VSVG used in the first generation EDV (as discussed in the Examples above), VSVGm-aCD3 does not have broad tropism – it preferentially targets T cells. $1\times10^5$ human T cells were transduced with $5\times10^{10}$ particles of EDV and $3\times10^{10}$ particles of AAV6 or Ark312. Ninety-six hours after transduction, the cells were analyzed by flow cytometry. When combined with AAV6 or Ark312 for transduction of activated human primary T cells, mVSVG-aCD3 EDV achieved similar *TRAC* CAR knock-in rates compared to the WT-VSVG EDV (Figures 2A-2B).

**Example 5 – The Anti-CD3 Antibody on the EDV Activated Naïve T cells**

**[0474]** The anti-CD3 antibody delivered by the second generation EDV (VSVGm-aCD3) was analyzed for activation of naïve T cells. Naïve T cells were treated with either 5 µl (low dose) or 25 µl (high dose) of concentrated EDVs and were compared to untransduced (UT) naïve T cells.

CD25 expression was used as a marker for T cell activation. The T cells were analyzed by flow cytometry 48 hours post transduction. As shown in Figure 2C, the VSVGm-aCD3 induced CD25 expression on T cells.

**Example 6 – *TRAC*-CAR T Cells were Toxic Against B Cells**

**[0475]** This Example demonstrates that CD19-targeting *TRAC*-CAR T cells (*TRAC*-1928z-1XX CAR T cells) were cytotoxic against the NALM6 B-ALL cell line. First, *TRAC*-1928z-1XX CAR T cells were generated by transducing activated T cells with WT-VSVG EDVs carrying Cas9/sg*TRAC* in combination with either AAV6 or Ark312 carrying an 1928z-1XX CAR HDRT. Then, NALM6 cells expressing luciferase were co-cultured with the transduced T cells at three effector to tumor cell (E:T) ratios: 1:1, 1:2, and 1:4. T cell cytotoxicity was determined by luminescence. As shown in Figure 2D, the *TRAC*-1928z-1XX CAR T cells caused toxicity in NALM6 cells.

**Example 7 – *In vivo* Editing of a T cell Genome Generates a CAR T Cell Capable of Killing Target Cells**

**[0476]** Next, the EDV and AAV combinations were assessed for generation of *TRAC*-targeted CAR T cells *in vivo*. A humanized mouse model was used as discussed in Example 1 and shown in Figure 3A. Fourteen days after human PBMC engraftment, the mice received IV injections of EDV and AAV combinations or PBS (control) as shown in Table 8 below. Two weeks after the mice received the EDV and AAV injections, the spleens of treated mice were harvested and analyzed by flow cytometry.

**Table 8: EDV and AAV Combinations Administered to Mice**

| **EDV carrying Cas9/sg*TRAC*** | | **AAV carrying 1928z-1XX-P2A-EGFRT HDRT** | | **Control** | **n** |
|---|---|---|---|---|---|
| *VSVG-WT* | *VSVGm-αCD3* | *AAV6* | *Ark312* | *PBS* | |
| $5\times10^{10}$ particles | n/a | $1\times10^{12}$ particles | n/a | n/a | 2 |
| n/a | $5\times10^{10}$ particles | $1\times10^{12}$ particles | n/a | n/a | 2 |
| $5\times10^{10}$ particles | n/a | n/a | $1\times10^{12}$ particles | n/a | 3 |

| | | | | | |
|---|---|---|---|---|---|
| n/a | $5\times10^{10}$ particles | n/a | $1\times10^{12}$ particles | n/a | 3 |
| n/a | n/a | n/a | n/a | yes | 2 |

**[0477]** CAR+ T cells were detected with all four combinations of EDVs and AAVs. WT-VSVG EDV and AAV6 produced about $2x10^5$ CAR T cells and VSVGm-aCD3 EDV and Ark312 produced about $3x10^6$ CAR T cells (Figure 3B). While both (1) VSVGm-aCD3 EDV and AAV6 and (2) WT-VSVG and Ark312 combinations resulted in more CAR T cells than WT-VSVG EDV and AAV6 combination (Figure 3B), the association of VSVGm-aCD3 EDV and Ark312 improvement was more than additive, suggesting a synergistic effect when the two vectors are combined. This observation was confirmed in both CD4 and CD8 T cells (Figure 3D). All combinations of EDVs and AAVs resulted in B cell aplasia (Figures 3B-3C). These results demonstrate *in vivo* editing of a T cell genome to generate a CAR T cell capable of killing target cells, which in this case, are B cells.

**Example 8 – VSVG targeted EDV and AAV6 EDVs can achieve targeted integrating in multiple primary and immortalized lines. Combining CD3-targeted EDV and Ark313 restrict engineering to T cells.**

**[0478]** A second major barrier to *in vivo* delivery is lack of specificity for the target cell population—in this case, human T cells. To further improve the selectivity of our EDV/AAV delivery system, we restricted the EDV specificity by incorporating a mutated VSVG with ablated affinity for the LDLR family of receptors (Nikolic, J. *et al. Nat Commun* 9, 1029 (2018).) and added an anti-CD3 (αCD3) single-chain variable fragment (scFv) that confers both targeting and activation (**Fig. 4a**). To confirm the specificity of our delivery method, we designed an assay where our dual EDV/AAV treatment leads to the specific integration of a promoter-less GFP in the first exon of the broadly expressed Clathrin A (*CLTA*) gene through HDR (Nguyen, D. N. *et al. Nat Biotechnol* 38, 44-49 (2020)). As *CLTA* is broadly expressed by all cell types, GFP expression indicates successful integration due to cell uptake of both the AAV and EDV. We used combinations of the original and evolved AAV to deliver a template for GFP knock-in at the *CLTA* locus with either VSVG or αCD3 pseudotyped EDVs to deliver *CLTA* targeting Cas9-RNP (**Fig. 4b**). Primary human T cells, NK cells, macrophages, HSCs and a panel of human B cell cancer

cell lines (NALM6, SupB15, JeKo1, Raji) were treated at a fixed MOI ($3 \times 10^5$ sgRNA/cell EDV, $5 \times 10^5$ vg/cell AAV) of EDV/AAV combinations for GFP-CLTA knock-in, and expression – indicative of correct integration – was assessed by flow cytometry (**Fig. 4b,c**).

**[0479]** As expected, VSVG/AAV6 provided broad delivery tropism, with significant $GFP^+$ populations across all tested cell lineages (**Fig. 4c, 5a,b**). Remarkably, Ark312 successfully abolished HSC targeting, and the combination of Ark312 with αCD3-EDV conferred selectivity for the human T cell compartment ($CD4^+$ and $CD8^+$), despite overall lowered efficiency when using αCD3-EDV (**Fig. 4c, 5a**). Since HSCs did not express the receptors to take up our EDV designs, we compared the efficacy of HDRT delivery using AAV6 and Ark312 after RNP electroporation and observed significant reduction of GFP-CLTA knock-in when using Ark312 (**Fig. 5c**), confirming that Ark312 is significantly de-targeted from human HSCs. Importantly, CAR expression in tumour cells after *in vivo* delivery could prevent cell surface expression of the CAR target and cause antigen-negative relapse. Therefore, it was encouraging that Ark312 reduced knock-in efficiency in all B cell cancer cell lines tested, and the combination of αCD3-EDV/Ark312 completely abolished cancer cell knock-in (**Fig. 4c, 5b**). The efficiencies of integration measured by flow cytometry were further confirmed by digital PCR (dPCR) using genomic DNA (**Fig. 5d**).

**[0480]** To assess the functional capacity of *in vivo*-generated CAR-T cells to control tumours, we employed NSG-MHC-I/II dKO mice into which we injected an aggressive leukaemia cell line, NALM6, followed by PBMCs three days later (**Fig. 6a**). EDV/AAV were injected one day post-PBMC injection (**Fig. 6a**), and as we previously demonstrated that the αCD3/Ark312 combination was the only one to achieve robust B cell aplasia, we proceeded with only this treatment (**Fig. 6b**). We selected a PBMC donor that alone had no impact on tumour growth and survival. Remarkably, after a single injection of the AAV/EDV, 5/9 mice achieved complete responses. All the mice that controlled the tumour were rechallenged with NALM6 ($5 \times 10^6$ cells) at D39 post EDV/AAV treatment and euthanized at D52 for organ harvest and analysis by flow cytometry (**Fig. 6d-g, 7a,b**). The mice treated with EDV/AAV were also able to control a rechallenge with NALM6, with no obvious tumour increase over a 2-week period post-rechallenge (**Fig. 6b**).

**[0481]** Analysis of the rechallenged mice demonstrated high numbers of total CAR-T cells in both the bone marrow and spleen and elimination of NALM6 cells in these organs (**Fig. 6d,e,**

**7a,b**). Of the total $CD45^+$ population in these organs, up to 40% of cells were $CAR^+$ after EDV/AAV delivery (**Fig. 6e, 7a**). Furthermore, we observed both $CD8^+$ and $CD4^+$ CAR-T cells in both bone marrow and spleen of the mice injected with EDV/AAV, with a more dominant $CD4^+$ population in bone marrow compared to spleen (**Fig. 6f**).

**[0482]** Together, these data demonstrate the first site-specific *in vivo* CAR-T cell generation using dually optimized AAV and EDV tools. This method can be employed to successfully reprogram T cells *in vivo* to mount an anti-tumour cell therapy response against an aggressive B cell cancer xenograft model.

**Example 9 – Ark315 Enables Efficient *In Vivo* T Cell Engineering**

**[0483]** Ark312 and Ark315 were compared for the ability to generate TRAC-CAR T cells in a xeno-GvHD free humanized mouse model (human PBMCs engrafted in MHCI/II dKO NSG mice). Two EDV/AAV combinations composed of αCD3 pseudotyped EDVs combined with Ark312 or Ark315 were tested. Five days post-injection of human PBMCs, an AAV was delivered with an EDV. The AAV carried a CD19-28z-1XX-P2A-EGFRT TRAC-HDRT. The EDV contained TRAC Cas9-RNP (Figure 8A). Two weeks post-vector injection, mice were euthanized and spleens harvested for flow cytometry.

**[0484]** The results show that both AAVs resulted in equivalent TRAC-CAR-T cell generation, averaging ~3% of all splenic T cells (Figures 8B and 8C). Both conditions resulted in systematic complete B cell aplasia (Figure 8C). These results demonstrate that combining αCD3-EDV with Ark312 or Ark315 offers efficient T cell targeting, which resulted in a deeper B cell aplasia.

**Example 10 – Combining EDV and AAV Can Generate TRAC-TCR T Cells**

**[0485]** In this Example, the combination of AAV and EDV was evaluated for the ability to generate targeted integration of a recombinant TCR into the TRAC locus. To test the system, an Ark312 AAV was generated to deliver a recombinant TCR specific for the NY-ESO-1 peptide to the TRAC locus (Figure 9A). Activated primary human T cells *in vitro* TRAC-targeted Cas9 EDVs were treated with an Ark312 delivering a TRAC-CAR HDRT ($3\times10^5$ sgRNA/cell EDV, $2\times10^5$ vg/cell AAV) and analyzed by flow cytometry the treated T cells 4 days later.

**[0486]** EDV alone led to TCR knock-out in >70% of both CD4 and CD8 T cells. Treatment with Ark312/EDV generated T cells binding to NEYSO dextramer (Figure 9B) reaching > 50% in CD4 T cells and >70% in CD8 T cells, indicating successful generation of TRAC-TCR-T cells by this method in vitro (Figures 9B and 9C).

**Overview for Examples 11-12 – Generation of TRAC-CAR T Cells Using Lipid Nanoparticles (LNPs)**

**[0487]** Examples 11-12 below relate to using AAVs and LNPs loaded with Cas9 nuclease to integrate a CAR transgene in the *TRAC* locus of T cells.

**Example 11 – Materials and Methods**

**[0488]** The materials and methods for Example 12 below are provided here in Example 11.

**[0489]** ***In vitro* knock-in (KI) in primary human T cells using mRNA LNP and AAV – Cell culture.** Primary human T cells (bulk) were cultured and activated using standard lab protocol. Briefly, T cells were cultured in T cell media supplemented with IL-7 and IL-15 and activated for two days. Before transfection, cells were washed with serum free medium and plated at 100,000 cells//well in 50 µL T cell medium without serum. Cells were immediately transfected with LNP co-encapsulating chemically modified Cas9 mRNA (TriLink BioTechnologies) and TRAC sgRNA together with AAV6 (AAV-U6/TRAC-1XX-EGFRt). Cells were transfected at a total RNA dose of 500 ng/well and 1E5/cell for AAV. 3-5 hours later, cells were supplemented with serum containing T cell media. After 5 days cells were collected and stained for FACS analysis.

**[0490]** **Lipid Nanoparticle (LNP) formulation.** LNPs were formulated using standard ethanol injection method. Ionizable lipids (SM-102, MC3, CL4H6, ssPalm-O-Phe, ALC-0315 and LP01), 18:1 Δ9-cis phosphoethanolamine (DOPE), cholesterol, and 14:0 PEG2000 phosphoethanolamine (C14-PEG2000) were dissolved in ethanol at a molar ratio of 50:30:20:2.5, respectively, to form the organic phase of the formulation. The aqueous phase was prepared by dissolving the corresponding mRNA in a 10 mM citrate buffer at pH 3 (Teknova, Hollister, CA, USA). Organic and aqueous phases were combined under vigorous mixing for 30 seconds and left for 10 minutes for assembly and particle formulation. LNPs were then washed twice with PBS and organic solvents were removed using Amicon® filters (Sigma-Aldrich). To rapidly navigate the chemical space of LNP chemical composition, design of experiments (DOEs) were used to reduce number

of experiments and identify hot spots. Using JMP software, a definitive screening design was used to identify the most critical factors contributing to efficient KI. These variations included type of ionizable lipid, type of phospholipid, and total lipid content (lipid/RNA ratio).

**[0491]** **LNP Characterization.** RNA encapsulation efficiency for LNPs was measured using a Quant-iT™ RiboGreen assay (Thermo Fisher Scientific). Each LNP sample was diluted 100-fold in two microcentrifuge tubes containing either 1X tris-EDTA (TE) buffer or 1X TE buffer supplemented with 1% (v/v) Triton X-100. The Triton X-100 samples were mixed thoroughly and allowed to incubate for 5 min to achieve lysis of LNPs. A standard curve was generated by diluting the corresponding mRNA used in the LNPs to concentrations ranging from 2.00 µg/mL to 31.3 ng/mL in 1X TE buffer. LNPs in TE buffer and LNPs in Triton X-100 were plated in quadruplicate, while the mRNA standards were plated in duplicate in black 96-well plates. Afterwards, the RiboGreen fluorescent detection reagent was added per manufacturer's instructions. The plate was then wrapped in aluminum foil and shook on a plate shaker at 200 rpm for 5 min. Afterwards, fluorescence intensity was read on a plate reader at an excitation wavelength of 490 nm and an emission wavelength of 530 nm. The hydrodynamic diameters and polydispersity indexes (PDIs) of the LNPs were measured using a Malvern Zetasizer instrument.

**Example 12 - Combining LNP and AAV Can Generate TRAC CAR T Cells**

**[0492]** To assess the efficacy of Cas9 and single- guide RNA (sgRNA) delivery using an LNP with an AAV to T cells *in vitro*, isolated and activated human primary T cells were treated with the TRAC gRNA LNP and TRAC-CAR-EGFRt AAV and quantified TRAC-knock-in efficiency by flow cytometry.

**[0493]** Co-treatment of T cells with LNP and AAV led to more than 15% TRAC knock-in (Figure 10B). The results demonstrate that LNPs can be used to integrate a transgene in the *TRAC* locus of T cells.

**Example 13 - In Vitro Genome Editing in Primary Human T Cells using EDV and AAV**

**[0494]** In this example, primary human T cells were cultured in X-VIVO™ 15 serum-free hematopoietic cell medium (Lonza) supplemented with 5% FBS (VWR), 50 uM 2-mercaptoethanol (Gibco), 10 mM N-acetyl L-cysteine (Sigma-Aldrich), and 300 U/mL

recombinant IL-2 (PeproTech). T cells were activated with anti-CD3/anti-CD28 Dynabeads (Gibco) at a ratio of 1:1 for 48 hours. After removing the beads, T cells were simultaneously transduced with a constant volume of VSVG-pseudotyped EDVs packaging TRAC-targeting Cas9 RNPs, and various MOIs of different AAV variants. The AAV variants packaged a transgene encoding an anti-CD19-1XX CAR and EGFRt between human specific TRAC homology arms. The AAV MOIs were calculated based on vector genome titers from qPCRs against the AAV2 ITRs that were present for every vector. Transductions were performed at high density (15,000 cells in 50 uL final), in serum-free medium, and in 96-well flat bottom plates before adding serum containing media (200 uL final) 12-16 hours later. Three days post-transfection, the cells were stained with anti-human Brilliant Violet 421 ™ (BV412) TCRa/b (BioLegend), anti-PE anti-human EGFR (BioLegend), and Zombie NIR Fixable Viability dye (BioLegend) in Cell Staining Buffer (BioLegend) and analyzed on an Attune NxT Flow Cytometer (Thermo Fisher Scientific) with an Autosampler. Data was analyzed in FlowJo and the HDR rate was calculated by dividing the percentage of EGFRt positive cells (knock-in) by the percentage of TCRa/b negative cells (knock-out).

**[0495]** Figure 11 presents data showing the indicated engineered AAV variants mediated TRAC-specific CAR knock-in in human T cells.

**[0496]** It is understood that the examples and embodiments described herein are for illustrative purposes only and that various modifications or changes in light thereof will be suggested to persons skilled in the art and are to be included within the spirit and purview of this application and scope of the appended claims. All publications, patents, and patent applications cited herein are hereby incorporated by reference in their entirety for all purposes.

**WHAT IS CLAIMED:**

1. A method of *in vivo* DNA insertion into a gene in cells in a mammal, the method comprising,

administering to the mammal:

(a) a first vector that delivers a polynucleotide-guided nuclease to a cell nuclei, and

(b) a second vector comprising a donor template polynucleotide,

wherein a guide polynucleotide for the polynucleotide-guided nuclease is present in the first vector, the second vector, or both the first and second vectors; and the guide polynucleotide is targeted to a portion of a gene in the cell genome that is selectively expressed in the cells, wherein the gene is any gene in the cell genome;

wherein presence of the first vector and the second vector in the cells in the mammal results in cleavage of the gene that is selectively expressed in the cells and integration of the donor template polynucleotide into the gene.

2. The method of claim 1, wherein the first vector comprises the polynucleotide-guided nuclease or a polynucleotide encoding the polynucleotide-guided nuclease.

3. The method of claim 1 or 2, wherein the first vector is an enveloped delivery vehicle (EDV), wherein the EDV optionally comprises a cell-specific binding molecule.

4. The method of claim 1 or 2, wherein the first vector is a lipid nanoparticle (LNP), wherein the LNP optionally comprises a cell-specific binding molecule.

5. The method of any one of claims 1-4, wherein the first vector and/or the second vector selectively targets one or more type of immune cells.

6. The method of claim 5, wherein the immune cells are T cells and the cell-specific binding molecule binds to CD3, CD4, CD5, CD7, CD8, CD28, 4-1BB ligand, T cell receptor (TCR) α constant chain, TCR β constant chain, or a major histocompatibility complex (MHC) carrying T cell receptor (TCR) specific peptide.

7. The method of claim 5, wherein the immune cells are B cells and the cell-specific binding molecule binds to CD19, CD20, BCMA, CD138, TACI, or CD22.

8. The method of claim 5, wherein the immune cells are NK cells and the cell-specific binding molecule binds to CD56, CD16, NKp46/NCR1, NCR2, or KIR.

9. The method of claim 5, wherein the immune cells are monocytes or macrophages and the cell-specific binding molecule binds to CD11b, CD68, CD14, CD33, or CD163.

10. The method of claim 5, wherein the immune cells are dendritic cells and the cell-specific binding molecule binds to CD11b, CD11c, XCR1, CD33, CD1c, or CD123.

11. The method of any one of claims 1-4, wherein the first vector and/or the second vector selectively targets stem cells.

12. The method of claim 11, wherein the stem cells are hematopoietic stem cells (HSCs) and the cell-specific binding molecule binds to CD34, CD117, CD49f, CD38, CD90, or EPCR.

13. The method of any one of claims 1-12, wherein the first vector further comprises a protein that catalyzes membrane fusion.

14. The method of any one of claims 1-13, wherein the first vector comprises the polynucleotide encoding the polynucleotide-guided nuclease and wherein the polynucleotide is an RNA molecule.

15. The method of any one of claims 1-14, wherein the second vector is a recombinant Adeno-associated virus (AAV) vector that has T-cell tropism.

16. The method of claim 15, wherein the AAV vector comprises an AAV capsid variant having reduced antibody-mediated neutralization compared to a wild-type capsid.

17. The method of any one of claims 1-16, wherein the guide polynucleotide is a guide RNA and the polynucleotide-guided nuclease is a CRISPR-Cas endonuclease.

18. The method of any one of claims 1-6 or 13-17, wherein the gene in the T cell genome is T cell receptor α constant (TRAC), T cell receptor β constant (TRBC), T cell receptor γ constant (TRGC), T cell receptor δ constant (TRDC), CD4, CD5, CD6, CD7, CD8a, CD8b, CD3e, CD247, CD27, CD28, interleukin-2R (IL-2R) α, IL-2R beta, killer cell lectin-like receptor (KLR) C1 (KLRC1), KLRF1, KLRG1, granzyme (GZM) A (GZMA), GZMB, GZMH, GZMK, Zap-70, lymphocyte-specific protein tyrosine kinase (LCK), linker for activation of T cells (LAT), IL-2 inducible T cell kinase (ITK), transcription factor 7 (TCF7), lymphoid enhancer binding factor 1 (LEF1), or FoxP3.

19. The method of any one of claims 1-6 or 13-18, wherein the guide polynucleotide targets the polynucleotide-guided nuclease to an exon of a TRAC gene.

20. The method of claim any one of claims 1-6 or 13-19, wherein the guide polynucleotide targets the polynucleotide-guided nuclease to exon 1 of the TRAC gene.

21. The method of any one of claims 1-6 or 13-18, wherein the guide polynucleotide targets the polynucleotide-guided nuclease to an intron of a TRAC gene.

22. The method of any one of claims 1-5, 7, or 13-17, wherein the gene in the B cell genome is CD19, CD20, CD22, CD138, BCMA, TACI, MS4A1, IGH, IGK, CD79A, or CD79B.

23. The method of any one of claims 1-5, 8, or 13-17, wherein the gene in the NK cell genome is NCAM1, FCGR3A, NCR1, NCR2, KLRC1, NKG2D, NKG7, KIR2DL1, KIR2DL2, KIR2DL3, KIR2DL4, KIR3DL1 or KIR3DL1.

24. The method of any one of claims 1-5, 9, or 13-17, wherein the gene in the monocyte or macrophage genome is CD11b, CD11c, CD14, CD33, CD163, CLEC7A, C1QA, C1QB, C1QC, or MSR1.

25. The method of any one of claims 1-5, 10, or 13-17, wherein the gene in the dendritic cell genome is CD1C, DCIR, CLEC10A, NDRG2, or TPM2.

26. The method of any one of claims 1-4 or 11-17, wherein the gene in the HSC genome is PTPRC, CD34, HBB, or RAG2.

27. The method of any one of claims 1-26, wherein the protein that catalyzes membrane fusion is a vesicular stomatitis virus glycoprotein G protein (VSVG) or a fusogenic variant thereof.

28. The method of claim 1, wherein

(a) the first vector is (i) an EDV optionally comprising a T cell-specific binding molecule and a protein that catalyzes membrane fusion, or (ii) an LNP optionally comprising a T cell-specific binding molecule; and

(b) the second vector is a recombinant Adeno-associated virus (AAV) vector having T cell tropism,

wherein the guide polynucleotide targets the polynucleotide-guided nuclease to exon 1 of a TRAC gene.

29. The method of claim 28, wherein the first vector is

(a) the EDV, wherein the EDV comprises the polynucleotide-guided nuclease, or

(b) the LNP, wherein the LNP comprises a polynucleotide encoding the polynucleotide-guided nuclease and wherein the polynucleotide is an RNA molecule.

30. The method of claim 29, wherein the donor template polynucleotide comprises a coding sequence for a polypeptide comprising a chimeric antigen receptor (CAR), a T cell receptor (TCR), or an HLA-independent T cell receptor (HIT).

31. The method of claim 29, wherein the donor template polynucleotide comprises a coding sequence for a polypeptide comprising (a) an extracellular target-binding domain, (b) a transmembrane domain, (c) a hinge domain, and (d) an intracellular signaling domain.

32. The method of claim 31, wherein the polypeptide extracellular target-binding domain binds to a CD19 polypeptide.

33. The method of claim 31 or 32, wherein the polypeptide transmembrane domain and the polypeptide hinge domain are derived from a CD28 polypeptide.

34. The method of any one of claims 31-33, wherein the polypeptide intracellular signaling domain comprises a modified CD3ζ polypeptide comprising (a) a native ITAM1, (b) a modified ITAM2 comprising two Tyr to Phe mutations, (c) a modified ITAM3 comprising two Tyr to Phe mutations, (d) a native BRS1, (e) a native BRS2, (f) a native BRS3, and (g) a co-stimulatory signaling region comprising a CD28 polypeptide.

35. The method of any one of claims 1-34, wherein the donor template polynucleotide is a homology-dependent repair template (HDRT) polynucleotide, homology-mediated end-joining template (HMEJT) polynucleotide, or a homology-independent targeted integration template (HITIT) polynucleotide.

36. The method of any one of claims 30-35, wherein after integration of the donor template polynucleotide into the gene, the coding sequence from the donor template polynucleotide is under control of endogenous promoter and/enhancer sequences.

37. A composition comprising:

(a) a first vector that delivers a polynucleotide-guided nuclease to a cell nuclei; and

(b) a second vector comprising a donor template polynucleotide,

wherein a guide polynucleotide for the polynucleotide-guided nuclease is present in the first vector, the second vector, or both the first and second vectors; and the guide polynucleotide is targeted to a portion of a gene in the cell genome that is selectively expressed in the cells; and

(c) a pharmaceutical carrier and/or a pharmaceutical excipient.

38. The composition of claim 37, wherein the first vector is an EDV or an LNP, wherein the EDV or the LNP optionally comprises a cell-specific binding molecule.

39. The composition of claim 37 or 38, wherein the cell-specific binding molecule is a T cell-specific binding molecule that binds to CD3, CD4, CD5, CD7, CD8, CD28,

4-1BB ligand, T cell receptor (TCR) α constant chain, TCR β constant chain, or a major histocompatibility complex (MHC) carrying T cell receptor (TCR) specific peptide.

40. The composition of claim 37 or 38, wherein the cell-specific binding molecule is a B cell-specific binding molecule that binds to to CD19, CD20, BCMA, CD138, TACI, or CD22.

41. The composition of claim 37 or 38, wherein the cell-specific binding molecule is an NK cell-specific binding molecule that binds to CD56, CD16, NKp46/NCR1, NCR2, or KIR.

42. The composition of claim 37 or 38, wherein the cell-specific binding molecule is a monocyte or macrophage-specific binding molecule that binds to CD11b, CD68, CD14, CD33, or CD163.

43. The composition of claim 37 or 38, wherein the cell-specific binding molecule is a dendritic cell-specific binding molecule that binds to CD11b, CD11c, XCR1, CD33, CD1c, or CD123.

44. The composition of claim 37 or 38, wherein the cell-specific binding molecule is an HSC-specific binding molecule that binds to CD34, CD117, CD49f, CD38, CD90, or EPCR.

45. The composition of any one of claims 37-44, wherein the first vector further comprises a protein that catalyzes membrane fusion.

46. The composition of any one of claims 37-45, wherein the second vector is a recombinant Adeno-associated virus (AAV) vector that has T-cell tropism.

47. The composition of claim 46, wherein the AAV vector comprises an AAV capsid variant having reduced antibody-mediated neutralization compared to a wild-type capsid.

48. The composition of any one of claims 37-47, wherein the guide polynucleotide is a guide RNA and the polynucleotide-guided nuclease is a CRISPR-Cas endonuclease.

49. The composition of any one of claims 37-39 or 45-48, wherein the guide polynucleotide targets the polynucleotide-guided nuclease to an exon of a TRAC gene.

50. The composition of any one of claims 37-39 or 45-49, wherein the guide polynucleotide targets the polynucleotide-guided nuclease to exon 1 of the TRAC gene.

51. The composition of any one or claims 45-50, wherein the protein that catalyzes membrane fusion is a vesicular stomatitis virus glycoprotein G protein (VSVG) or a fusogenic variant thereof.

52. The composition of any one of claims 37-51, wherein the donor template polynucleotide comprises a coding sequence for a polypeptide comprising a chimeric antigen receptor (CAR), a T cell receptor (TCR), or an HLA-independent T cell receptor (HIT).

53. The composition of any one of claims 37-51, wherein the donor template polynucleotide comprises a coding sequence for a polypeptide comprising (a) an extracellular target-binding domain, (b) a transmembrane domain, (c) a hinge domain, and (d) an intracellular signaling domain.

54. The composition of claim 53, wherein the polypeptide extracellular target-binding domain binds to a CD19 polypeptide.

55. The composition of claim 53 or 54, wherein the polypeptide transmembrane domain and the polypeptide hinge domain are derived from a CD28 polypeptide.

56. The composition of any one of claims 53-55, wherein the polypeptide intracellular signaling domain comprises a modified CD3ζ polypeptide comprising (a) a native ITAM1, (b) a modified ITAM2 comprising two Tyr to Phe mutations, (c) a modified ITAM3 comprising two Tyr to Phe mutations, (d) a native BRS1, (e) a native BRS2, (f) a native BRS3, and (g) a co-stimulatory signaling region comprising a CD28 polypeptide.

57. The composition of any one of claims 37-56, wherein the donor template polynucleotide is a homology-dependent repair template (HDRT) polynucleotide, homology-mediated end-joining template (HMEJT) polynucleotide, or a homology-independent targeted integration template (HITIT) polynucleotide.