(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
24 October 2024 (24.10.2024)

WIPO | PCT

(10) International Publication Number
WO 2024/220574 A1

(51) International Patent Classification:
*C12N 15/86* (2006.01)

(21) International Application Number:
PCT/US2024/025039

(22) International Filing Date:
17 April 2024 (17.04.2024)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

| | | |
|---|---|---|
| 63/460,304 | 18 April 2023 (18.04.2023) | US |
| 63/466,716 | 15 May 2023 (15.05.2023) | US |
| 63/522,713 | 22 June 2023 (22.06.2023) | US |


(71) Applicant: **SANA BIOTECHNOLOGY, INC.** [US/US]; 188 East Blaine Street, Suite 400, Seattle, Washington 98102 (US).

(72) Inventors: **STOPPATO, Matteo**; 188 East Blaine Street, Suite 400, Seattle, Washington 98102 (US). **TRUDEAU, Kyle Marvin**; 188 East Blaine Street, Suite 400, Seattle, Washington 98102 (US). **FRYE, Zachary P.**; 188 East Blaine Street, Suite 400, Seattle, Washington 98102 (US). **SHAH, Jagesh Vijaykumar**; 188 East Blaine Street, Suite 400, Seattle, Washington 98102 (US). **REILLY, Jack**; 188 East Blaine Street, Suite 400, Seattle, Washington 98102 (US).

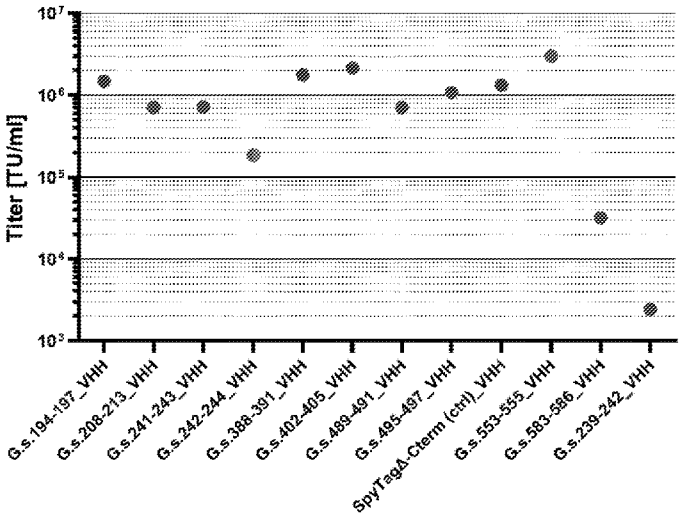
(74) Agent: **SCHMIDT, Jeffrey W.** et al.; Morrison & Foerster LLP, 12531 High Bluff Drive, Suite 100, San Diego, California 92130-2040 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ,

(54) Title: UNIVERSAL PROTEIN G FUSOGENS AND ADAPTER SYSTEMS THEREOF AND RELATED LIPID PARTICLES AND USES

FIG. 1B




(57) Abstract: Provided herein are lipid particles, such as lenti viral particles, that incorporate or are pseudotyped with paramyxovirus envelope attachment glycoproteins as fused to at least one component of a universal adapter system, and in some aspects also a fusion (F) protein such as a NiV-F protein or a biologically active portion or variant thereof. The present disclosure also provides for preparation of the lipid particles, such as lentiviral particles, comprising at least one component of the universal adapter system, as well as methods for preparing and using the lipid particles, such as lentiviral particles.

WO 2024/220574 A1

*[Continued on next page]*

RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, ME, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

**Published:**

— *with international search report (Art. 21(3))*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*
— *with sequence listing part of description (Rule 5.2(a))*

# UNIVERSAL PROTEIN G FUSOGENS AND ADAPTER SYSTEMS THEREOF AND RELATED LIPID PARTICLES AND USES

## Cross-Reference to Related Applications

**[0001]** This application claims priority to U.S. Provisional Patent Application No. 63/460,304, filed April 18, 2023, entitled "UNIVERSAL PROTEIN G FUSOGENS AND ADAPTER SYSTEMS THEREOF AND RELATED LIPID PARTICLES AND USES", to U.S. Provisional Patent Application No. 63/466,716, filed May 15, 2023, entitled "UNIVERSAL PROTEIN G FUSOGENS AND ADAPTER SYSTEMS THEREOF AND RELATED LIPID PARTICLES AND USES", and to U.S. Provisional Patent Application No. 63/522,713, filed June 22, 2023, entitled "UNIVERSAL PROTEIN G FUSOGENS AND ADAPTER SYSTEMS THEREOF AND RELATED LIPID PARTICLES AND USES", the contents of each of which are herein incorporated by reference in their entirety for all purposes.

## Reference to an Electronic Sequence Listing

**[0002]** The present application is being filed along with a Sequence Listing in electronic format. The Sequence Listing is provided as a file entitled 18615-20085.40.XML created April 17, 2024 which is 752,788 bytes in size. The information in the electronic format of the Sequence Listing is incorporated by reference in its entirety.

## Field

**[0003]** The present disclosure relates to lipid particles, such as lentiviral particles, that incorporate or are pseudotyped with paramyxovirus envelope attachment glycoproteins as fused to at least one component of a universal adapter system, and in some aspects also a fusion (F) protein such as a NiV-F protein or a biologically active portion or variant thereof. The present disclosure also provides for preparation of the lipid particles, such as lentiviral particles, comprising at least one component of the universal adapter system, as well as methods for preparing and using the lipid particles, such as lentiviral particles.

## Background

**[0004]** Lipid particles, including viral-based particles like virus-like particles and viral vectors such as lentiviral particles, are commonly used for delivery of exogenous agents to cells. For various particles, such as lentiviral vector particles, the host range can be altered by pseudotyping with a heterologous envelope protein or a modified envelope protein which further comprises a targeting moiety. The

efficient preparation and production of particles with certain pseudotyped envelope proteins with a targeting moiety may not always be efficient, such as due to effects of the envelope protein and/or targeting moiety on low titer of the produced lentiviral vector particles. Improved lipid particles, including virus-like particles and viral vectors, that can be produced with a higher titer and with efficient transduction efficiency of target cells are needed. The provided disclosure addresses this need.

Summary

**[0005]** Provided herein is a lipid particle, comprising (a) a paramyxovirus envelope attachment protein attached to a targeting moiety via a universal adapter system, said system comprising a tag polypeptide component and a catcher polypeptide component, wherein the paramyxovirus envelope attachment protein is fused to one of the components, and the targeting moiety is fused to the other of the components; and (b) at least one paramyxovirus fusion (F) protein; wherein the protein in (a) and (b) are exposed on the outside of the lipid bilayer. In some of any of the provided embodiments, the tag polypeptide component and catcher polypeptide component are covalently bound.

**[0006]** Provided herein is a lipid particle, comprising (a) a paramyxovirus envelope attachment protein fused to a tag polypeptide component or fused to a catcher polypeptide component of a universal adapter system; and (b) at least one paramyxovirus fusion (F) protein; wherein the protein in (a) and (b) are exposed on the outside of the lipid bilayer.

**[0007]** In some of any of the provided embodiments, the paramyxovirus envelope attachment protein is fused to the tag polypeptide component of the universal adapter system. In some of any of the provided embodiments, the paramyxovirus envelope attachment protein is fused to the catcher polypeptide component of the universal adapter system. In some of any of the provided embodiments, the paramyxovirus envelope attachment protein is fused to the component of the universal adapter system at the C terminus of the paramyxovirus envelope attachment protein. In some of any of the provided embodiments, the paramyxovirus envelope attachment protein is fused to the component of the universal adapter system by insertion of the component within a peptide loop of the paramyxovirus envelope attachment protein.

**[0008]** In some of any of the provided embodiments, the peptide loop of the paramyxovirus envelope attachment protein is selected from the group consisting of: (i) amino acid residues 194-197; (ii) amino acid residues 208-213; (iii) amino acid residues 241-243; (iv) amino acid residues 242-243; (v) amino acid residues 388-391; (vi) amino acid residues 402-405; (vii) amino acid residues 489-491; and (viii) amino acid residues 495-497; optionally wherein the peptide loop is 388-391, 489-491, or 495-497.

**[0009]** In some of any of the provided embodiments, the fusion is direct or indirect via flexible linker, optionally a peptide linker. In some of any of the provided embodiments, the peptide linker is a

poly-Glycine-Serine (G4S) linker, optionally wherein the linker is set forth in any one of SEQ ID NOs. 7-10.

**[0010]** In some of any of the provided embodiments, the universal adapter system comprises a tag component and a catcher component from a SpyTag/SpyCatcher system; a SnoopTag/SnoopCatcher system; a TEFCA tag/catcher system, or a DogTag/DogCatcher system. In some of any of the provided embodiments, the universal adapter system comprises a SpyTag tag component and a SpyCatcher catcher component selected from the group comprising: (i) SpyTagΔ and SpyCatcherΔ (e.g., SEQ ID NOs 11 and 16); or (ii) SpyTag003 and SpyCatcher003 (e.g., SEQ ID NOs 12 and 17). In some of any of the provided embodiments, the universal adapter system comprises: (i) a SnoopTag tag component (e.g., SEQ ID NO. 13) and a SnoopCatcher catcher component (e.g., SEQ ID NO. 18); (ii) a TEFCA tag component (e.g., SEQ ID NO. 14) and a TEFCACatcher catcher component (e.g., SEQ ID NO. 19); or (iii) a DogTag tag component (e.g., SEQ ID NO. 15) and a DogCatcher catcher component (e.g., SEQ ID NO. 20).

**[0011]** In some of any of the provided embodiments, the targeting moiety is specific for a target molecule expressed on the surface of a target cell. In some of any of the provided embodiments, the targeting moiety is selected from the group consisting of an antibody or antigen-binding fragment, an engineered binding domain, a nanobody, a DARPin, an Aptamer, an Affimer, an Affibody, a Knottin, an Avimer, a Monobody, an Anticalin, a Fynomer, and a targeting peptide. In some of any of the provided embodiments, the targeting moiety is selected from the group consisting of a single domain antibody or a single chain variable fragment (scFv). In some of any of the provided embodiments the single domain antibody is a VHH.

**[0012]** In some of any of the provided embodiments, the paramyxovirus envelope attachment protein is an envelope attachment protein from a Nipah virus, Hendra virus, or Measles virus, or is a variant or biologically active portion thereof of any of the foregoing. In some of any of the provided embodiments, the paramyxovirus envelope attachment protein is a variant Nipah G protein (NiV-G) that is a variant or a biologically active portion of a wild-type NiV-G. In some of any of the provided embodiments, the variant NiV-G tis truncated by up to 40 contiguous amino acids at or near the N-terminus of the wild-type NiV-G set forth in SEQ ID NO:5. In some of any of the provided embodiments, the variant NiV-G has a truncation of amino acids 2-34 of the wild-type NiV-G set forth in SEQ ID NO:5. In some of any of the provided embodiments, the paramyxovirus envelope attachment protein comprises one or more mutations that reduces native tropism relative to the wild-type paramyxovirus envelope attachment protein not comprising the one or more mutations. In some of any of the provided embodiments, the variant NiV-G exhibits reduced binding to Ephrin B2 or Ephrin B3. In some of any of the provided embodiments, the variant NiV-G comprises: one or more amino acid substitutions corresponding to amino acid substitutions selected from the group consisting of E501A, W504A, Q530A and E533A with reference to numbering set forth in SEQ ID NO:5. In some of any of

the provided embodiments, the variant NiV-G comprises amino acid substitutions E501A, W504A, Q530A and E533A with reference to numbering set forth in SEQ ID NO:5. In some of any of the provided embodiments, the variant NiV-G has the amino acid sequence set forth in SEQ ID NO: 6 or an amino acid sequence having at or about 80%, at least at or about 81 %, at least at or about 82%, at least at or about 83%, at or about 84%, at least at or about 85%, at least at or about 86%, or at least at or about 87%, at least at or about 88%, or at least at or about 89%, at least at or about 90%, at least at or about 91 %, at least at or about 92%, at least at or about 93%, at least at or about 94%, at least at or about 95%, at or about 96%, at least at or about 97%, at least at or about 98%, or at least at or about 99% sequence identity to SEQ ID NO:6. In some of any of the provided embodiments, the variant NiV-G has the amino acid sequence set forth in SEQ ID NO:6.

**[0013]** In some of any of the provided embodiments, the paramyxovirus F protein is an F0 precursor or is a proteolytically cleaved form thereof comprising F1 and F2 subunits. In some of any of the provided embodiments, the proteolytically cleaved form is a cathepsin L cleavage product. In some of any of the provided embodiments, the at least one paramyxovirus fusion (F) protein is a wild-type Nipah F protein (NiV-F) or is a functionally active variant or a biologically active portion thereof.

**[0014]** In some of any of the provided embodiments, the NiV-F protein or the biologically active portion is a truncated NiV-F that is truncated by at least 22 amino acids or at least 20 amino acids, at or near the C-terminus of wild-type NiV-F set forth in SEQ ID NO:1. In some of any of the provided embodiments, the NiV-F protein or the biologically active portion thereof has a 22 amino acid truncation at or near the C-terminus of the wild-type NiV-F protein. In some of any of the provided embodiments, the NiV-F protein comprises a deletion in it cytoplasmic tail and lacks amino acid residues 525-546 of SEQ ID NO:1. In some of any of the provided embodiments, the NiV-F protein or the biologically active portion thereof has the sequence set forth in SEQ ID NO: 3 or 4, or a sequence of amino acids that exhibits at least at or about 80%, 85%, 90% or 95% sequence identity to the sequence set forth in SEQ ID NO: 3 or 4.

**[0015]** In some of any of the provided embodiments, the lipid particle is a vector, optionally a viral vector. In some of any of the provided embodiments, the vector is a lentiviral vector. In some of any of the provided embodiments, the lipid particle further comprises an exogenous agent for delivery to the target cell. In some of any of the provided embodiments, the exogenous agent is present in the lumen. In some of any of the provided embodiments, the exogenous agent is a protein or a nucleic acid, optionally wherein the nucleic acid is a DNA or RNA. In some of any of the provided embodiments, the exogenous agent is a nucleic acid encoding a cargo for delivery to the target cell. In some of any of the provided embodiments, the exogenous agent is or encodes a therapeutic agent, a diagnostic agent or a genome-modifying enzyme. In some of any of the provided embodiments, the exogenous agent encodes a membrane protein, optionally wherein the membrane protein is an antigen receptor for targeting cells

expressed by or associated with a disease or condition. In some of any of the provided embodiments, the membrane protein is a chimeric antigen receptor (CAR).

**[0016]** Provided herein is a polynucleotide comprising a nucleic acid encoding the paramyxovirus attachment protein of any of the provided lipid particles, wherein the paramyxovirus envelope attachment protein is fused to a tag polypeptide component or a catcher polypeptide component of a universal adapter system. Provided herein is a polypeptide comprising the paramyxovirus envelope attachment protein of any of the provided lipid particles, wherein the paramyxovirus envelope attachment protein is fused to a tag polypeptide component or a catcher polypeptide component of a universal adapter system. In some embodiments, the fusion is by insertion of the component into a loop region of the paramyxovirus envelope attachment protein as described.

**[0017]** In some of any of the provided embodiments, the nucleic acid sequence is a first nucleic acid sequence and the polynucleotide further comprises a second nucleic acid sequence encoding a paramyxovirus fusion (F) protein of any of the provided lipid particles. In some of any of the provided embodiments, the polynucleotide comprises an IRES or a sequence encoding a linking peptide between the first and second nucleic acid sequences. In some embodiments, the linking peptide is a self-cleaving peptide or a peptide that causes ribosome skipping. In some embodiments, the linking peptide is a T2A peptide. In some of any of the provided embodiments, the polynucleotide further comprises at least one promoter that is operatively linked to control expression of the nucleic acid. In some embodiments, the promoter controls expression of the first nucleic acid sequence and the second nucleic acid sequence.

**[0018]** Provided herein is a vector, comprising any of the provided polynucleotides. In some of any of the provided embodiments, the vector is a mammalian vector, viral vector or artificial chromosome, optionally wherein the artificial chromosome is a bacterial artificial chromosome (BAC). Provided herein is a plasmid, comprising any of the provided polynucleotides. In some of any of the provided embodiments, the plasmid further comprises one or more nucleic acids encoding proteins for lentivirus production.

**[0019]** Provided herein is a cell comprising any of the provided nucleotides, vectors, or plasmids.

**[0020]** Provided herein is a method of making a lipid particle, comprising: a) providing a cell that comprises any of the provided nucleotides, vectors, or plasmids; b) culturing the cell under conditions that allow for production of a lipid particle, and c) separating, enriching, or purifying the targeted lipid particle from the cell, thereby making the targeted lipid particle. In some embodiments, the cell is a producer host cell.

**[0021]** Provided herein is a method of making a pseudotyped lentiviral vector, comprising: a) providing a producer cell that comprises a lentiviral viral nucleic acid(s), and any of the provided nucleotides, vectors, or plasmids; b) culturing the cell under conditions that allow for production of the lentiviral vector, and c) separating, enriching, or purifying the lentiviral vector from the cell, thereby making the pseudotyped lentiviral vector.

**[0022]** Provided herein is a producer cell comprising any of the provided nucleotides, vectors, or plasmids. In some of any of the provided embodiments, the producer cell further comprises nucleic acid encoding a paramyxovirus F protein or a biologically active portion thereof. In some of any of the provided embodiments, the cell further comprises a viral nucleic acid. In some embodiments, the viral nucleic acid is a lentiviral nucleic acid. Provided herein is a lipid particle produced by any of the methods.

**[0023]** Provided herein is a composition comprising a plurality of any of the provided lipid particles. In some of any of the provided embodiments, the composition further comprises a pharmaceutically acceptable carrier.

**[0024]** Provided herein is a combination, comprising (a) any of the provided lipid particles, wherein the paramyxovirus envelope attachment protein is fused to a tag polypeptide component or a catcher polypeptide component of a universal adapter system, and (b) a targeting moiety fused to the other component of the universal adapter system.

**[0025]** In some of any of the provided embodiments, the universal adapter system comprises a tag component and a catcher component from a SpyTag/SpyCatcher system; a SnoopTag/SnoopCatcher system; a TEFCA tag/catcher system, or a DogTag/DogCatcher system. In some of any of the provided embodiments, the universal adapter system comprises a SpyTag tag component and a SpyCatcher catcher component selected from the group comprising: (i) SpyTagΔ and SpyCatcherΔ (e.g., SEQ ID NOs 11 and 16); or (ii) SpyTag003 and SpyCatcher003 (e.g., SEQ ID NOs 12 and 17). In some of any of the provided embodiments, the universal adapter system comprises: (i) a SnoopTag tag component (e.g., SEQ ID NO. 13) and a SnoopCatcher catcher component (e.g., SEQ ID NO. 18); (ii) a TEFCA tag component (e.g., SEQ ID NO. 14) and a TEFCACatcher catcher component (e.g., SEQ ID NO. 19); or (iii) a DogTag tag component (e.g., SEQ ID NO. 15) and a DogCatcher catcher component (e.g., SEQ ID NO. 20).

**[0026]** In some of any of the provided embodiments, the targeting moiety is specific for a target molecule expressed on the surface of a target cell. In some of any of the provided embodiments, the targeting moiety is selected from the group consisting of an antibody or antigen-binding fragment, an engineered binding domain, a nanobody, a DARPin, an Aptamer, an Affimer, an Affibody, a Knottin, an Avimer, a Monobody, an Anticalin, a Fynomer, and a targeting peptide. In some of any of the provided embodiments, the targeting moiety is specific for a target molecule expressed on the surface of a target cell. In some of any of the provided embodiments, the targeting moiety is selected from the group consisting of a single domain antibody or a single chain variable fragment (scFv). In some of any of the provided embodiments, the single domain antibody is a VHH.

**[0027]** Provided herein is a method of making a lipid particle comprising a paramyxovirus envelope attachment protein attached to a targeting moiety via a universal adapter system, said system comprising a tag polypeptide component and a catcher polypeptide component, wherein the paramyxovirus envelope

attachment protein is fused to one of the components, and the targeting moiety is fused to the other of the components; and at least one paramyxovirus fusion (F) protein, the method comprising: (a) providing a cell that comprises any of the provided lipid particles; (b) culturing the cell under condition that allow for production of the lipid particle, (c) separating, enriching, or purifying the lipid particle from the cell, thereby making the lipid particle; (d) mixing the lipid particle with a targeting moiety fused to the other component of the universal adapter system; and (e) removing unbound targeting moiety from the lipid particles to collect lipid particles with the paramyxovirus envelope attachment protein attached to the targeting moiety. In some of any of the provided embodiments, the method further comprises concentrating the collected lipid particles.

**[0028]** Provided herein is a lipid particle produced by any of the provided methods. Provided herein is a kit comprising any of the provided combinations. Also provided herein is an article of manufacture comprising any of the provided kits.

**[0029]** Provided herein is a method of selective transduction, the method comprising contacting a target cell with any of the provided lipid particles. Also provided herein is a method of selective transduction, the method comprising exposing a composition comprising a target cell with (1) any of the provided lipid particles, wherein the lipid particle comprises a paramyxovirus envelope attachment protein fused to a tag polypeptide component or a catcher polypeptide component of a universal adapter system; and (2) a targeting moiety fused to the other component of the universal adapter system. In some of any of the provided embodiments, the method is performed in vitro or ex vivo. In some of any of the provided embodiments, the method is performed in vivo.

**[0030]** Provided herein is a method of selective transduction of a target cell in a subject, the method comprising administering any of the provided lipid particles to a subject. Also provided herein is a method of selective transduction of a target cell in a subject, the method comprising: (a) administering any of the provided lipid particles to a subject, wherein the lipid particle comprises a paramyxovirus envelope attachment protein fused to a tag polypeptide component or a catcher polypeptide component of a universal adapter system; and (b) administering a targeting moiety fused to the other component of the universal adapter system to the subject.

**[0031]** In some of any of the provided embodiments, the universal adapter system comprises a tag component and a catcher component from a SpyTag/SpyCatcher system; a SnoopTag/SnoopCatcher system; a TEFCA tag/catcher system, or a DogTag/DogCatcher system. In some of any of the provided embodiments, the universal adapter system comprises a SpyTag tag component and a SpyCatcher catcher component selected from the group comprising: (i) SpyTagΔ and SpyCatcherΔ (e.g., SEQ ID NOs 11 and 16); or (ii) SpyTag003 and SpyCatcher003 (e.g., SEQ ID NOs 12 and 17). In some of any of the provided embodiments, wherein the universal adapter system comprises: (i) a SnoopTag tag component (e.g., SEQ ID NO. 13) and a SnoopCatcher catcher component (e.g., SEQ ID NO. 18); (ii) a TEFCA tag component (e.g., SEQ ID NO. 14) and a TEFCACatcher catcher component (e.g., SEQ ID NO. 19); or

(iii) a DogTag tag component (e.g., SEQ ID NO. 15) and a DogCatcher catcher component (e.g., SEQ ID NO. 20). . In some of any of the provided embodiments, the targeting moiety is specific for a target molecule expressed on the surface of the target cell. In some of any of the provided embodiments, the targeting moiety is selected from the group consisting of an antibody or antigen-binding fragment, an engineered binding domain, a nanobody, a DARPin, an Aptamer, an Affimer, an Affibody, a Knottin, an Avimer, a Monobody, an Anticalin, a Fynomer, and a targeting peptide. In some of any of the provided embodiments, the targeting moiety is selected from the group consisting of a single domain antibody or a single chain variable fragment (scFv). In some of any of the provided embodiments, the single domain antibody is a VHH.

**[0032]** In some of any of the provided embodiments, the contacting is in vivo in a subject. In some of any of the provided embodiments, the method treats a disease or condition in the subject by the delivery of the exogenous agent. Provided herein is a method of treating a disease or disorder in a subject (e.g., a human subject), the method comprising administering to the subject any of the provided lipid particles or compositions.

**[0033]** Provided herein is a paramyxovirus envelope attachment protein with a targeting moiety attached via a universal adapter system, said system comprising a tag polypeptide component and a catcher polypeptide component, wherein the paramyxovirus envelope attachment protein is fused to one of the components, and the targeting moiety is fused to the other of the components.

**[0034]** Also provided herein is a paramyxovirus envelope attachment protein comprising one component of with a targeting moiety attached via a universal adapter system, said system comprising a tag polypeptide component and a catcher polypeptide component, wherein the paramyxovirus envelope attachment protein is fused to the one component.

**[0035]** In some of any of the provided embodiments, the component fused to the paramyxovirus envelope attachment protein is linked to the C-terminus of the paramyxovirus envelope attachment protein. In some of any of the provided embodiments, the component fused to the paramyxovirus envelope attachment protein is inserted into a loop region of the paramyxovirus envelope attachment protein. In some of any of the provided embodiments, the loop is selected from the group consisting of: (i) amino acid residues 194-197; (ii) amino acid residues 208-213; (iii) amino acid residues 241-243; (iv) amino acid residues 242-243; (v) amino acid residues 388-391; (vi) amino acid residues 402-405; (vii) amino acid residues 489-491; and (viii) amino acid residues 495-497; optionally wherein the peptide loop is 388-391, 489-491, or 495-497. In some of any of the provided embodiments, wherein the paramyxovirus envelope attachment protein is from Nipah. In some of any of the provided embodiments, the paramyxovirus envelope attachment protein is a NiV-G or is a biologically active variant or truncated form thereof. In some of any of the provided embodiments, wherein the targeting moiety is selected from the group consisting of an antibody or antigen-binding fragment, an engineered binding domain, a

nanobody, a DARPin, an Aptamer, an Affimer, an Affibody, a Knottin, an Avimer, a Monobody, an Anticalin, a Fynomer, and a targeting peptide.

Brief Description of the Drawings

**[0036]** Titer of exemplary lentiviral vector (LV) produced by mixing vector comprising SpyTagΔ (e.g., a tag component of a universal adapter system) at various insertion sites located in the NiV-G protein with a cognate CD8 retargeted scFv (**FIG. 1A**) or VHH (**FIG. 1B**) binder fused to SpyCatcher retargeted to CD8. **FIG. 1C** shows G/p24 area as a function of G expression and **FIG. 1D** the corresponding occupancy for two exemplary insertion sites of a component of a universal adapter system

**[0037]** **FIG. 2** depicts transduction efficiency of pseudotyped vector retargeted covalently to the binder via the universal adapter system (CD8 retargeted universal LV). Titers of CD8 retargeted universal LV are shown in **FIG. 3A**. Titers for viral vectors pseudotyped with exemplary configurations of components of a universal adapter system (e.g., Spy/Tag) are shown in **FIG. 3B. FIG. 4** shows G/p24 area as a function of G expression for pseudotyped LVs with CD8 retargeted SpyTagΔ as the universal adaptor.

**[0038]** NiV-G expression as compared with exemplary NiV-G + binder-Catcher is shown as a Western blot in **FIG. 5A**. Occupancy is similarly shown in **FIG. 5B** as percentage of G with binder (NiV-G + binder).

**[0039]** Transduction efficiency of the produced pseudotyped lentiviral vector with binder (i.e., vector with binder comprising one of two exemplary VHH constructs or one of four exemplary scFv constructs) was assessed for titer on transduced cells as shown in **FIG. 5C**. Similarly, transduction efficiency of the produced pseudotyped lentiviral vector with binder (i.e., vector with binder comprising one of three exemplars) was assessed for titer on transduced cells as shown in **FIG. 5D.**

**[0040]** **FIG. 6A** depicts titers for various loop insertion sites. Titer as a function of G protein occupancy is shown in **FIG. 6B**. **FIG. 6C** shows a positive correlation between occupancy and titer.

**[0041]** **FIG. 7** depicts a sequence of an exemplary NiV-G envelope attachment glycoprotein (SEQ ID NO:6) with exemplary regions into which a component of a universal adaptor system can be inserted, shown (underlined and bolded).

**[0042]** **FIG. 8** depicts transduction of two different HEK 293 cell lines with pseudotyped retargeted universal LV.

**[0043]** **FIG. 9** shows methods of preparation for lentiviral vector and binder (e.g., universal adapter system).

**[0044]** **FIG. 10** shows both Western blot and exoview analysis of G protein per lentiviral particle.

**[0045]** **FIG. 11A** shows titers for lentiviral vector engineered with two alternative adapter-modified Gs ($G_{ADAPTER1}$+ $G_{ADAPTER2}$) as well as a $G_{WT}$ control vector. Potency for the same lentiviral vector

engineered with two alternative adapter-modified Gs ($G_{ADAPTER1}$+ $G_{ADAPTER2}$) as well as the $G_{WT}$ control vector is shown in **FIG. 11B.**

**[0046]** **FIG. 12** shows titers for lentiviral vectors pseudotyped with a NiV-G protein having a DogTag within a loop region (388-391, 402-045, 495-497, and/or 553-555) and a NiV-F protein.

Detailed Description

**[0047]** Provided herein are lipid particle comprising a paramyxovirus envelope attachment protein. For instance, provided herein are lipid particles containing a lipid bilayer enclosing a lumen or cavity and a paramyxovirus envelope attachment protein, wherein the paramyxovirus envelope attachment protein is fused to a tag polypeptide component or is fused to a catcher polypeptide component of a universal adapter system. In some embodiments, the paramyxovirus envelope attachment protein is attached to a targeting moiety via a universal adapter system, said system comprising a tag polypeptide component and a catcher polypeptide component, wherein the paramyxovirus envelope attachment protein is fused to one of the components, and the targeting moiety is fused to the other of the components. In particular embodiments, the lipid particles can be a virus-like particle, a virus, or a viral vector, such as a lentiviral vector.

**[0048]** In some embodiments, any of the provided lipid particles also contains a viral fusion (F) protein molecule or a biologically active portion thereof embedded in the lipid bilayer. In some embodiments, the F protein is from a Paramyxovirus, or is a biologically active portion thereof or is a variant or mutant thereof. In particular embodiments, the F protein is from a Nipah (NiV) virus.

**[0049]** In naturally occurring paramyxoviruses, the fusion (F) and attachment (G, H, or HN) glycoproteins mediate cellular entry of paramyxovirus, such as Nipah virus. In some embodiments, the combination of an F protein, such as a NiV-F protein, and NiV-G protein as provided herein is able to mediate cellular entry of a provided lipid particle (e.g. lentiviral vector).

**[0050]** The F protein, such as Nipah Virus F protein, also known as NiV-F, is a class I fusion protein that has structural and functional features in common with fusion proteins of many families (e.g., HIV-1 gp41 or influenza virus hemagglutinin [HA]), such as an ectodomain with a hydrophobic fusion peptide and two heptad repeat regions (White JM et al. 2008. Crit Rev Biochem Mol Biol 43:189–219). F proteins are synthesized as inactive precursors $F_0$ and are activated by proteolytic cleavage into the two disulfide-linked subunits $F_1$ and $F_2$ (Moll M. et al. 2004. J. Virol. 78(18): 9705-9712).

**[0051]** G proteins are attachment proteins of paramyxovirus such as henipavirus (e.g. Nipah virus or Hendra virus) that are type II transmembrane glycoproteins containing an N-terminal cytoplasmic tail, a transmembrane domain, an extracellular stalk, and a globular head (Liu, Q. et al. 2015. Journal of Virology, 89(3):1838-1850). The Nipah virus attachment protein, NiV-G, recognizes the receptors EphrinB2 and EphrinB3. Binding of the receptor to NiV-G triggers a series of conformational changes

that eventually lead to the triggering of NiV-F, which exposes the fusion peptide of NiV-F, allowing another series of conformational changes that lead to virus-cell membrane fusion (Stone J.A. et al. 2016. J Virol. 90(23): 10762-10773). EphrinB2 was previously identified as the primary NiV receptor (Negrete et al., 2005), as well as ephrinB3 as an alternate receptor (Negrete et al., 2006). In fact, wild-type NiV-G has an high affinity for ephrinB2 and B3, with affinity binding constants (Kd) in the picomolar range (Negrete et al., 2006) (Kd=0.06 nM and 0.58 nM for cell surface expressed ephrinB2 and B3, respectively).

**[0052]** In some embodiments, the paramyxovirus envelope attachment protein may be further linked to a targeting moiety via a component of a universal adapter system, including targeting moieties such as an antigen binding domain, to facilitate specific targeting of the lipid particle to a target molecule for fusion with a desired target cell.

**[0053]** Thus, the provided embodiments, the paramyxovirus envelope attachment protein may be re-targeted to any desired cell type for specific targeting of a lipid particle (e.g. lentiviral vector) and, in some cases, specific delivery to a target cell of a transgene or heterologous protein (e.g., exogenous agent) contained therein.

**[0054]** The efficiency of transduction of lipid particles can be improved by engineering mutations in one or both of NiV-F and NiV-G. Several such mutations have been previously described (see, e.g., Lee at al., 2011, Trends in Microbiology). This could be useful, for example, for maintaining the specificity and picomolar affinity of NiV-G for ephrinB2 and/or B3. Additionally, mutations in NiV-G that completely abrogate ephrinB2 and B3 binding, but that do not impact the association of this NiV-G with NiV-F, have been identified (Aguilar, et al. *J Biol Chem.* 2009;284(3):1628-1635.; Weise et al. *J Virol.* 2010;84(15):7634-764; Negrete et al.. *J Virol.* 2007;81(19):10804-10814; Negrete et al. *PLoS Pathog.* 2006; Guillaume et al., *J. Virol* 2006, 80 (15) 7546-7554 In some cases, methods to improve targeting of lipid particles can be achieved by fusion of a binding molecule with a paramyxovirus envelope attachment protein (e.g. NiV-G, including a Niv-G with mutations to abrogate Ephrin B2 and Ephrin B3 binding). This could allow for altered G protein tropism allowing for targeting of other desired cell types that are not ephrinB2+ through the addition of the binding molecule directed against a different cell surface molecule via a component of a universal adapter system. Thus, in provided aspects, a paramyxovirus envelope attachment protein (e.g., NiV-G) may further contain a mutation to reduce or abrogate binding to Ephrin B2 and/B3. In some embodiments, the mutations can include one or more of mutations E501A, W504A, Q530A and E533A, with reference to numbering of wild-type NiV-G set forth in SEQ ID NO:5.

**[0055]** The provided lipid particles, such as lentiviral vectors, containing a paramyxovirus envelope attachment protein exhibit advantages over available envelope-pseudotyped particles. For instance, VSV-G is the most common envelope glycoprotein used for pseudotyping but its broad tropism is often not ideal or desirable for specific target cell delivery, such as is desired for gene therapy or exogenous

protein delivery. Further, although alternative envelope proteins may exhibit reduced tropism or may be amenable to linkage to a binding domain for redirected targeting to a desired target cells, the titer of a preparation of lentiviral vectors containing such envelope proteins may be too low to allow for efficient transduction. Thus, alternative approaches are needed, such as an approach described herein using components of a universal adapter system.

**[0056]** It is found herein that certain paramyxovirus envelope attachment protein when pseudotyped on a lentiviral vector with a component of a universal adapter system exhibit high titers. Further, in provided aspects, the titers are superior to other paramyxovirus envelope attachment protein, including those fused directly to a targeting moiety in the absence of any component of a universal adapter system. In particular, it is found herein that certain paramyxovirus envelope attachment protein when pseudotyped on a lentiviral vector with a component of a universal adapter system as provided herein result in improved titer of a lentiviral preparation. Finally, combining the paramyxovirus envelope attachment protein when pseudotyped on a lentiviral vector with a component of a universal adapter system with an F protein (e.g. a NiV-F or a biologically active portion or variant thereof) retains high fusion activity of the lentiviral vector, and in some cases also delivery of a transgene or other exogenous agent, to a target cell.

**[0057]** Also provided are lipid particles, such as targeted lipid particles, additionally containing one or more exogenous agents, such as for delivery of a diagnostic or therapeutic agent to cells, including following in vivo administration to a subject. Also provided herein are methods and uses of the lipid particles, such in diagnostic and therapeutic methods. Also provided are polynucleotides, methods for engineering, preparing, and producing the lipid non-cell particles, compositions containing the particles, and kits and devices containing and for using, producing and administering the particles.

**[0058]** All publications, including patent documents, scientific articles and databases, referred to in this application are incorporated by reference in their entirety for all purposes to the same extent as if each individual publication were individually incorporated by reference. If a definition set forth herein is contrary to or otherwise inconsistent with a definition set forth in the patents, applications, published applications and other publications that are herein incorporated by reference, the definition set forth herein prevails over the definition that is incorporated herein by reference.

**[0059]** The section headings used herein are for organizational purposes only and are not to be construed as limiting the subject matter described.

## DEFINITIONS

**[0060]** Unless defined otherwise, all terms of art, notations and other technical and scientific terms or terminology used herein are intended to have the same meaning as is commonly understood by one of ordinary skill in the art to which the claimed subject matter pertains. In some cases, terms with commonly understood meanings are defined herein for clarity and/or for ready reference, and the

inclusion of such definitions herein should not necessarily be construed to represent a substantial difference over what is generally understood in the art. Unless indicated otherwise, abbreviations and symbols for chemical and biochemical names is per IUPAC-IUB nomenclature. Unless indicated otherwise, all numerical ranges are inclusive of the values defining the range as well as all integer values in-between.

**[0061]** As used herein, the articles "a" and "an" refer to one or to more than one (i.e. to at least one) of the grammatical object of the article. By way of example, "an element" means one element or more than one element.

**[0062]** As used herein, the term "about" will be understood by persons of ordinary skill in the art and will vary to some extent on the context in which it is used. As used herein, "about" when referring to a measurable value such as an amount, a temporal duration, and the like, is meant to encompass variations of ±20% or ±10%, more preferably ±5%, even more preferably ±1%, and still more preferably ±0.1% from the specified value, as such variations are appropriate to perform the disclosed methods.

**[0063]** As used herein, "lipid particle" refers to any biological or synthetic particle that contains a bilayer of amphipathic lipids enclosing a lumen or cavity. Typically a lipid particle does not contain a nucleus. Such lipid particles include, but are not limited to, viral particles (e.g. lentiviral particles), virus-like particles, viral vectors (e.g., lentiviral vectors) exosomes, enucleated cells, various vesicles, such as a microvesicle, a membrane vesicle, an extracellular membrane vesicle, a plasma membrane vesicle, a giant plasma membrane vesicle, an apoptotic body, a mitoparticle, a pyrenocyte, or a lysosome. In some embodiments, a lipid particle can be a fusosome. In some embodiments, the lipid particle is not a platelet. In some embodiments, the fusosome is derived from a source cell. A lipid particle also may include an exogenous agent or a nucleic acid encoding an exogenous agent, which may be present in the lumen of the lipid particle.

**[0064]** The terms "viral vector particle" and "viral vector" are used interchangeably herein and refer to a vector for transfer of an exogenous agent (e.g. non-viral or exogenous nucleic acid) into a recipient or target cell and that contains one or more viral structural proteins in addition to at least one non-structural viral genomic component or functional fragment thereof (i.e., a polymerase, an integrase, a protease or other non-structural component). The viral vector thus contains the exogenous agent, such as heterologous nucleic acid that includes non-viral coding sequences, to be transferred into a cell. Examples of viral vectors are retroviral vectors, such as lentiviral vectors.

**[0065]** The term "retroviral vector" refers to a viral vector that contains retroviral nucleic acid or is derived from a retrovirus. A retroviral vector particle includes the following components: a vector genome (retrovirus nucleic acid), a nucleocapsid encapsidating the nucleic acid, and a membrane envelope surrounding the nucleocapsid. Typically, a retroviral vector contains sufficient retroviral genetic information to allow packaging of an RNA genome, in the presence of packaging components, into a viral particle capable of infecting a target cell. Infection of the target cell may include reverse

transcription and integration into the target cell genome. A retroviral vector may be a recombinant retroviral vector that is replication defective and lacks genes essential for replication, such as a functional gag-pol and/or env gene and/or other genes essential for replication. A retroviral vector also may be a self-inactivating (SIN) vector.

**[0066]** As used herein, a "lentiviral vector" or LV refers to a viral vector that contains lentiviral nucleic acid or is derived from a lentivirus. A lentiviral vector particle includes the following components: a vector genome (lentivirus nucleic acid), a nucleocapsid encapsidating the nucleic acid, and a membrane surrounding the nucleocapsid. Typically, a lentiviral vector contains sufficient lentiviral genetic information to allow packaging of an RNA genome, in the presence of packaging components, into a viral particle capable of infecting a target cell. Infection of the target cell may include reverse transcription and integration into the target cell genome. A lentiviral vector may be a recombinant lentiviral vector that is replication defective and lacks genes essential for replication, such as a functional gag-pol and/or env gene and/or other genes essential for replication. A lentiviral vector also may be a self-inactivating (SIN) vector.

**[0067]** As used herein, a "retroviral nucleic acid," refers to a nucleic acid containing at least the minimal sequence requirements for packaging into a retroviral vector, alone or in combination with a helper cell, helper virus, or helper plasmid. In the case of "lentiviral nucleic acid" the nucleic acid refers to at least the minimal sequence requirements for packaging into a lentiviral vector, alone or in combination with a helper cell, helper virus, or helper plasmid. In some embodiments, the viral nucleic acid comprises one or more of (e.g., all of) a 5' LTR (e.g., to promote integration), U3 (e.g., to activate viral genomic RNA transcription), R (e.g., a Tat-binding region), U5, a 3' LTR (e.g., to promote integration), a packaging site (e.g., psi (Ψ)), RRE (e.g., to bind to Rev and promote nuclear export). The viral nucleic acid can comprise RNA (e.g., when part of a virion) or DNA (e.g., when being introduced into a source cell or after reverse transcription in a recipient cell). In some embodiments, the viral nucleic acid is packaged using a helper cell, helper virus, or helper plasmid which comprises one or more of (e.g., all of) gag, pol, and env.

**[0068]** As used herein, "fusosome" refers to a lipid particle containing a bilayer of amphipathic lipids enclosing a lumen or cavity and a fusogen that interacts with the amphipathic lipid bilayer. In some embodiments, the fusosome is a membrane enclosed preparation. In some embodiments, the fusosome is derived from a source cell. A fusosome also may include an exogenous agent or a nucleic acid encoding an exogenous agent, which may be present in the lumen of the fusosome.

**[0069]** As used herein, "fusosome composition" refers to a composition comprising one or more fusosomes.

**[0070]** As used herein, "fusogen" refers to an agent or molecule that creates an interaction between two membrane enclosed lumens. In embodiments, the fusogen facilitates fusion of the membranes. In other embodiments, the fusogen creates a connection, e.g., a pore, between two lumens (e.g., a lumen of a

retroviral vector and a cytoplasm of a target cell). In some embodiments, the fusogen comprises a complex of two or more proteins, e.g., wherein neither protein has fusogenic activity alone. In some embodiments, the fusogen comprises a targeting domain. Examples of fusogens include paramyxovirus F and G proteins such as those from Nipah Virus (NiV) and biologically active portions or variants thereof including any as described.

**[0071]** As used herein, a "re-targeted fusogen," such as a re-targeted G protein, refers to a fusogen that comprises a targeting moiety having a sequence that is not part of the naturally-occurring form of the fusogen in which the targeting moiety targets or binds a molecule on a desired cell type. In embodiments, the fusogen comprises a different targeting moiety relative to the targeting moiety in the naturally-occurring form of the fusogen. In embodiments, the naturally-occurring form of the fusogen lacks a targeting domain, and the re-targeted fusogen comprises a targeting moiety that is absent from the naturally-occurring form of the fusogen. In embodiments, the fusogen is modified to comprise a targeting moiety. In some such embodiments, the attachment of the targeting moiety to a fusogen (e.g. G protein) may be directly or indirectly via a linker, such as a peptide linker. In embodiments, the fusogen comprises one or more sequence alterations outside of the targeting moiety relative to the naturally-occurring form of the fusogen, e.g., in a transmembrane domain, fusogenically active domain, or cytoplasmic domain.

**[0072]** As used herein, a "target cell" refers to a cell of a type to which it is desired that a lipid particle, such as a targeted lipid particle, delivers an exogenous agent. In embodiments, a target cell is a cell of a specific tissue type or class, e.g., an immune effector cell, e.g., a T cell. In some embodiments, a target cell is a diseased cell, e.g., a cancer cell. In some embodiments, the fusogen, e.g., re-targeted fusogen leads to preferential delivery of the exogenous agent to a target cell compared to a non-target cell.

**[0073]** As used herein a "non-target cell" refers to a cell of a type to which it is not desired that a lipid particle delivers an exogenous agent. In some embodiments, a non-target cell is a cell of a specific tissue type or class. In some embodiments, a non-target cell is a non-diseased cell, e.g., a non-cancerous cell. In some embodiments, the fusogen, e.g., re-targeted fusogen leads to lower delivery of the exogenous agent to a non-target cell compared to a target cell.

**[0074]** As used herein a "biologically active portion," such as with reference to a protein such as a G protein or an F protein, refers to a portion of the protein that exhibits or retains an activity or property of the full-length of the protein. For example, a biologically active portion of an F protein retains fusogenic activity in conjunction with the G protein when each are embedded in a lipid bilayer. A biologically active portion of the G protein retains fusogenic activity in conjunction with an F protein when each is embedded in a lipid bilayer. The retained activity can include 10%-150% or more of the activity of a full-length or wild-type F protein or G protein. Examples of biologically active portions of F and G

proteins include proteins with truncations of the cytoplasmic domain, such as any of the described variant NiV-F with a truncated cytoplasmic tail.

**[0075]** As used herein, "percent (%) amino acid sequence identity" and "homology" with respect to a peptide, polypeptide or antibody sequence are defined as the percentage of amino acid residues in a candidate sequence that are identical with the amino acid residues in the specific peptide or polypeptide sequence, after aligning the sequences and introducing gaps, if necessary, to achieve the maximum percent sequence identity, and not considering any conservative substitutions as part of the sequence identity. Alignment for purposes of determining percent amino acid sequence identity can be achieved in various ways that are within the skill in the art, for instance, using publicly available computer software such as BLAST, BLAST-2, ALIGN or MEGALIGN (DNASTAR) software. Those skilled in the art can determine appropriate parameters for measuring alignment, including any algorithms needed to achieve maximal alignment over the full length of the sequences being compared.

**[0076]** An amino acid substitution may include but are not limited to the replacement of one amino acid in a polypeptide with another amino acid. Exemplary substitutions are shown in Table 1. Amino acid substitutions may be introduced into an antibody of interest and the products screened for a desired activity, for example, retained/improved binding.

**Table 1**

| **Original Residue** | **Exemplary Substitutions** |
|---|---|
| Ala (A) | Val; Leu; Ile |
| Arg (R) | Lys; Gln; Asn |
| Asn (N) | Gln; His; Asp, Lys; Arg |
| Asp (D) | Glu; Asn |
| Cys (C) | Ser; Ala |
| Gln (Q) | Asn; Glu |
| Glu (E) | Asp; Gln |
| Gly (G) | Ala |
| His (H) | Asn; Gln; Lys; Arg |
| Ile (I) | Leu; Val; Met; Ala; Phe; Norleucine |
| Leu (L) | Norleucine; Ile; Val; Met; Ala; Phe |
| Lys (K) | Arg; Gln; Asn |
| Met (M) | Leu; Phe; Ile |
| Phe (F) | Trp; Leu; Val; Ile; Ala; Tyr |
| Pro (P) | Ala |
| Ser (S) | Thr |
| Thr (T) | Val; Ser |
| Trp (W) | Tyr; Phe |
| Tyr (Y) | Trp; Phe; Thr; Ser |
| Val (V) | Ile; Leu; Met; Phe; Ala; Norleucine |

**[0077]** Amino acids may be grouped according to common side-chain properties:

(1) hydrophobic: Norleucine, Met, Ala, Val, Leu, Ile;

(2) neutral hydrophilic: Cys, Ser, Thr, Asn, Gln;

(3) acidic: Asp, Glu;

(4) basic: His, Lys, Arg;

(5) residues that influence chain orientation: Gly, Pro;

(6) aromatic: Trp, Tyr, Phe.

**[0078]** Non-conservative substitutions will entail exchanging a member of one of these classes for another class.

**[0079]** The term, "corresponding to" with reference to positions of a protein, such as recitation that nucleotides or amino acid positions "correspond to" nucleotides or amino acid positions in a disclosed sequence, such as set forth in the Sequence listing, refers to nucleotides or amino acid positions identified upon alignment with the disclosed sequence based on structural sequence alignment or using a standard alignment algorithm, such as the GAP algorithm. For example, corresponding residues of a similar sequence (e.g. fragment or species variant) can be determined by alignment to a reference sequence by structural alignment methods. By aligning the sequences, one skilled in the art can identify corresponding residues, for example, using conserved and identical amino acid residues as guides.

**[0080]** The term "isolated" as used herein refers to a molecule that has been separated from at least some of the components with which it is typically found in nature or produced. For example, a polypeptide is referred to as "isolated" when it is separated from at least some of the components of the cell in which it was produced. Where a polypeptide is secreted by a cell after expression, physically separating the supernatant containing the polypeptide from the cell that produced it is considered to be "isolating" the polypeptide. Similarly, a polynucleotide is referred to as "isolated" when it is not part of the larger polynucleotide (such as, for example, genomic DNA or mitochondrial DNA, in the case of a DNA polynucleotide) in which it is typically found in nature, or is separated from at least some of the components of the cell in which it was produced, for example, in the case of an RNA polynucleotide. Thus, a DNA polynucleotide that is contained in a vector inside a host cell may be referred to as "isolated".

**[0081]** The term "effective amount" as used herein means an amount of a pharmaceutical composition which is sufficient to significantly and positively modify the symptoms and/or conditions to be treated (e.g., provide a positive clinical response). The effective amount of an active ingredient for use in a pharmaceutical composition will vary with the particular condition being treated, the severity of the condition, the duration of treatment, the nature of concurrent therapy, the particular active ingredient(s) being employed, the particular pharmaceutically-acceptable excipient(s) and/or carrier(s) utilized, and like factors with the knowledge and expertise of the attending physician.

**[0082]** An "exogenous agent" as used herein with reference to a lipid particle, such as a viral vector, refers to an agent that is neither comprised by nor encoded in the corresponding wild-type virus or fusosome made from a corresponding wild-type source cell. In some embodiments, the exogenous agent

does not naturally exist, such as a protein or nucleic acid that has a sequence that is altered (e.g., by insertion, deletion, or substitution) relative to a naturally occurring protein. In some embodiments, the exogenous agent does not naturally exist in the source cell. In some embodiments, the exogenous agent exists naturally in the source cell but is exogenous to the virus. In some embodiments, the exogenous agent does not naturally exist in the recipient cell. In some embodiments, the exogenous agent exists naturally in the recipient cell, but is not present at a desired level or at a desired time. In some embodiments, the exogenous agent comprises RNA or protein.

**[0083]** As used herein, a "promoter" refers to a cis- regulatory DNA sequence that, when operably linked to a gene coding sequence, drives transcription of the gene. The promoter may comprise a transcription factor binding sites. In some embodiments, a promoter works in concert with one or more enhancers which are distal to the gene.

**[0084]** As used herein, a composition refers to any mixture of two or more products, substances, or compounds, including cells. It may be a solution, a suspension, liquid, powder, a paste, aqueous, non-aqueous or any combination thereof.

**[0085]** As used herein, the term "pharmaceutically acceptable" refers to a material, such as a carrier or diluent, which does not abrogate the biological activity or properties of the compound, and is relatively nontoxic, i.e., the material may be administered to an individual without causing undesirable biological effects or interacting in a deleterious manner with any of the components of the composition in which it is contained.

**[0086]** As used herein, the term "pharmaceutical composition" refers to a mixture of at least one compound of the invention with other chemical components, such as carriers, stabilizers, diluents, dispersing agents, suspending agents, thickening agents, and/or excipients. The pharmaceutical composition facilitates administration of the compound to an organism. Multiple techniques of administering a compound exist in the art including, but not limited to, intravenous, oral, aerosol, parenteral, ophthalmic, pulmonary and topical administration.

**[0087]** A "disease" or "disorder" as used herein refers to a condition where treatment is needed and/or desired.

**[0088]** As used herein, the terms "treat," "treating," or "treatment" refer to ameliorating a disease or disorder, e.g., slowing or arresting or reducing the development of the disease or disorder or reducing at least one of the clinical symptoms thereof. For purposes of this disclosure, ameliorating a disease or disorder can include obtaining a beneficial or desired clinical result that includes, but is not limited to, any one or more of: alleviation of one or more symptoms, diminishment of extent of disease, preventing or delaying spread (for example, metastasis, for example metastasis to the lung or to the lymph node) of disease, preventing or delaying recurrence of disease, delay or slowing of disease progression, amelioration of the disease state, inhibiting the disease or progression of the disease, inhibiting or slowing the disease or its progression, arresting its development, and remission (whether partial or total).

**[0089]** The terms "individual" and "subject" are used interchangeably herein to refer to an animal; for example a mammal. The term patient includes human and veterinary subjects. In some embodiments, methods of treating mammals, including, but not limited to, humans, rodents, simians, felines, canines, equines, bovines, porcines, ovines, caprines, mammalian laboratory animals, mammalian farm animals, mammalian sport animals, and mammalian pets, are provided. The subject can be male or female and can be any suitable age, including infant, juvenile, adolescent, adult, and geriatric subjects. In some examples, an "individual" or "subject" refers to an individual or subject in need of treatment for a disease or disorder. In some embodiments, the subject to receive the treatment can be a patient, designating the fact that the subject has been identified as having a disorder of relevance to the treatment, or being at adequate risk of contracting the disorder. In particular embodiments, the subject is a human, such as a human patient.

## I. UNIVERSAL ADAPTER RETARGETED LIPID PARTICLE SYSTEM

**[0090]** Provided herein is a lipid particle, said lipid particle comprising a paramyxovirus envelope attachment protein attached to a targeting moiety via a universal adapter system. In some embodiments, the universal adapter system comprises two components: a tag polypeptide component and a catcher polypeptide component. In some embodiments, the paramyxovirus envelope attachment protein is fused to one of the components, and the targeting moiety is fused to the other of the components. For instance, in one embodiments, the paramyxovirus envelope attachment protein is fused to the tag component, while the targeting moiety is fused to the catcher component of the same universal adapter system. In another embodiment, a paramyxovirus envelope attachment protein is fused to one component of a universal adapter system (e.g., either of a tag or catcher component). In some embodiments, the lipid particle further comprises an F protein. In some embodiments, the paramyxovirus envelope attachment protein and paramyxovirus F protein are exposed on the outside of the lipid bilayer.

### A. Universal Adapter System

**[0091]** Provided herein are lipid particles which comprise components of a universal adapter system. In some embodiments, universal adapter systems are a coupling mechanism for irreversible protein-protein ligation and/or covalent bonding. Universal adapter systems such as disclosed herein comprise a tag component and a catcher component, which are capable of interaction with one another. In some embodiments, the tag component of a universal adapter system is a short, unfolded peptide that can be genetically fused to exposed positions in target proteins. Similarly, the catcher component can also be fused to reporter proteins such as GFP, and to epitope or purification tags. In some embodiment, a paramyxovirus envelope protein is fused to one component of a universal adapter system (e.g., a tag). In further embodiments, the targeting moiety is fused to the other component of a universal adapter system (e.g., a catcher).

**[0092]** For instance, the SpyCatcher-SpyTag system was developed as a method for protein ligation via a universal adapter. It is based on a modified domain from a *Streptococcus pyogenes* surface protein (SpyCatcher), which recognizes a cognate 13-amino-acid peptide (SpyTag). Upon recognition, the two form a covalent isopeptide bond between the side chains of a lysine in SpyCatcher and an aspartate in SpyTag. An additional (orthogonal) system called SnoopTag-SnoopCatcher has been developed from an *S. pneumoniae* pilin that can be combined with a catcher and tag (e.g.,SpyCatcher-SpyTag) to produce protein fusions with multiple components (Hatlem et al., Int. J. Mol. Sci. 2019, 20(9), 2129).

**[0093]** In some embodiments, the paramyxovirus envelope attachment protein is fused to a tag polypeptide component as set forth in **Table 2A.** In some embodiments, the paramyxovirus envelope attachment protein is fused to a catcher polypeptide component of a universal adapter system as set forth in **Table 2A.** In some embodiments, the targeting moiety is fused to a tag polypeptide component as set forth in **Table 2A.** In some embodiments, the targeting moiety is fused to a catcher polypeptide component of a universal adapter system as set forth in **Table 2A.** In some embodiment, the paramyxovirus envelope attachment protein is fused to a tag polypeptide component as set forth in **Table 2A** and the targeting moiety is fused to a catcher polypeptide component of a universal adapter system as set forth in **Table 2A.** In some embodiments, the paramyxovirus envelope attachment protein is fused to a catcher polypeptide component of a universal adapter system as set forth in **Table 2A** and the targeting moiety is fused to a tag polypeptide component as set forth in **Table 2A.**

**[0094]** In some embodiments, the paramyxovirus envelope attachment protein is fused to a tag polypeptide component as set forth in SEQ ID NO: 11, and the targeting moiety is fused to a catcher polypeptide component as set forth in SEQ ID NO: 16. In some embodiments, the paramyxovirus envelope attachment protein is fused to a catcher polypeptide component as set forth in SEQ ID NO: 16, and the targeting moiety is fused to a tag polypeptide component as set forth in SEQ ID NO: 11.

**[0095]** In some embodiments, the paramyxovirus envelope attachment protein is fused to a tag polypeptide component as set forth in SEQ ID NO: 12, and the targeting moiety is fused to a catcher polypeptide component as set forth in SEQ ID NO: 16. In some embodiments, the paramyxovirus envelope attachment protein is fused to a catcher polypeptide component as set forth in SEQ ID NO: 16, and the targeting moiety is fused to a tag polypeptide component as set forth in SEQ ID NO: 12.

**[0096]** In some embodiments, the paramyxovirus envelope attachment protein is fused to a tag polypeptide component as set forth in SEQ ID NO: 11, and the targeting moiety is fused to a catcher polypeptide component as set forth in SEQ ID NO: 17. In some embodiments, the paramyxovirus envelope attachment protein is fused to a catcher polypeptide component as set forth in SEQ ID NO: 17, and the targeting moiety is fused to a tag polypeptide component as set forth in SEQ ID NO: 11.

**[0097]** In some embodiments, the paramyxovirus envelope attachment protein is fused to a tag polypeptide component as set forth in SEQ ID NO: 12, and the targeting moiety is fused to a catcher polypeptide component as set forth in SEQ ID NO: 17. In some embodiments, the paramyxovirus

envelope attachment protein is fused to a catcher polypeptide component as set forth in SEQ ID NO: 17, and the targeting moiety is fused to a tag polypeptide component as set forth in SEQ ID NO: 12.

**[0098]** In some embodiments, the paramyxovirus envelope attachment protein is fused to a tag polypeptide component as set forth in SEQ ID NO: 13, and the targeting moiety is fused to a catcher polypeptide component as set forth in SEQ ID NO: 18. In some embodiments, the paramyxovirus envelope attachment protein is fused to a catcher polypeptide component as set forth in SEQ ID NO: 18, and the targeting moiety is fused to a tag polypeptide component as set forth in SEQ ID NO: 13.

**[0099]** In some embodiments, the paramyxovirus envelope attachment protein is fused to a tag polypeptide component as set forth in SEQ ID NO: 14, and the targeting moiety is fused to a catcher polypeptide component as set forth in SEQ ID NO: 19. In some embodiments, the paramyxovirus envelope attachment protein is fused to a catcher polypeptide component as set forth in SEQ ID NO: 19, and the targeting moiety is fused to a tag polypeptide component as set forth in SEQ ID NO: 14.

**[0100]** In some embodiments, the paramyxovirus envelope attachment protein is fused to a tag polypeptide component as set forth in SEQ ID NO: 15, and the targeting moiety is fused to a catcher polypeptide component as set forth in SEQ ID NO: 20. In some embodiments, the paramyxovirus envelope attachment protein is fused to a catcher polypeptide component as set forth in SEQ ID NO: 20, and the targeting moiety is fused to a tag polypeptide component as set forth in SEQ ID NO: 15.

**[0101]** In some embodiments, the paramyxovirus envelope attachment protein is fused to a tag polypeptide component as set forth in SEQ ID NO: 511, and the targeting moiety is fused to a catcher polypeptide component as set forth in SEQ ID NO: 512. In some embodiments, the paramyxovirus envelope attachment protein is fused to a catcher polypeptide component as set forth in SEQ ID NO: 512, and the targeting moiety is fused to a tag polypeptide component as set forth in SEQ ID NO: 511.

**Table 2A: Exemplary Components of a Universal Adapter System**

| Name | Sequence | SEQ ID NO |
|---|---|---|
| Universal Adapter System – Tag Components | | |
| SpyTagΔ | AHIVMVDAYKPTK | 11 |
| SpyTag003 | RGVPHIVMVDAYKRYK | 12 |
| SnoopTag | KLGDIEFIKVNK | 13 |
| TEFCA Tag | TEFCA | 14 |
| DogTag | DIPATYEFTDGKHYITNEPIPPK | 15 |
| Universal Adapter System – Catcher Components | | |

| | | |
|---|---|---|
| SpyCatcherΔ | DIPATYEFTDGKHYITNEPIPPK | 16 |
| SpyCatcher003 | DSATHIKFSKRDEDGKELAGAT MELRDSSGKTISTWISDGQVKD FYLYPGKYTFVETAAPDGYEVA TAITFTVNEQGQVTVNG | 17 |
| SnoopCatcher | VTTLSGLSGEQGPSGDMTTEED SATHIKFSKRDEDGRELAGATME LRDSSGKTISTWISDGHVKDFYL YPGKYTFVETAAPDGYEVATPIEF TVNEDGQVTVDGEATEGDAHT | 18 |
| TEFCACatcher | KPLRGAVFSLQKQHPDYPDIYG AIDQNGTYQNVRTGEDGKLTFK NLSDGKYRLFENSEPAGYKPVQ NKPIVAFQIVNGEVRDVTSIVPQ DIPATYEFTNGKHYITNEPIPPK | 19 |
| DogCatcher | AGELIHMVTLDKTGKKSFGICIVRGE VKDSPNTKTTGIFIKGIVPDSPAHL CGRLKVGDRILSLNGKDVRNSTEQ AVIDLIKEADFKIELEIQTFDK | 20 |

**[0102]** In some embodiments, universal adapter systems are a coupling mechanism for irreversible protein-protein ligation and/or covalent bonding. In some embodiments, universal adapter systems are a coupling mechanism for reversible protein-protein ligation and/or non-covalent bonding. Universal adapter systems such as disclosed herein may comprise biotin and/or streptavidin, which are capable of interaction with one another. Universal adapter systems such as disclosed herein comprise inteins, which are capable of interaction with one another.

**[0103]** Inteins are protein introns that autocatalytically splice from host polypeptides to generate a functional protein. Inteins can also function as mobile genetic elements that are expressed as intervening sequences within host proteins. Intein-containing proteins are expressed as precursor polypeptides, with the host protein sequence preceding the intein known as the N-extein and the sequence following the intein known as the C-extein. Inteins then undergo post-translational autocatalytic excision while forming the spliced (ligated) exteins, which constitute the mature host protein. The splicing component of inteins comprises around 150 amino acids that fold into a compact disc-shaped structure that brings the two exteins into proximity (Lennon et al., Current Biology, Vol 27, Issue 6, 2017).

**[0104]** For example, expressed protein ligation (EPL) is one exemplary approach that introduces intein-mediated C-terminal protein modifications. GyrA intein from *Mycobacterium xenopi* (Mxe), also known as "minimal" intein is comprised of 198 amino acid residues. In some embodiments, a modified full-length GyrA intein is first fused to the C terminus of the envelope attachment protein through recombinant protein expression. After intein-catalyzed N-S acyl shift, the GyrA intein is cleaved off from the attachment protein by the addition of a thiol reagent, such as Mesna, resulting in a C-terminus thioester bond. Cys-functionalized targeting moiety sequences are then conjugated to the attachment

protein. This forms a stable peptide bond with the attachment protein C-terminus, and the subsequent S-N acyl shift generates a new thiol group that forms a disulfide bond with its other Ct counterpart, further stabilizing the Ct conjugation area.

**[0105]** In some embodiments, the intein of a universal adapter system is a short, unfolded peptide that can be genetically fused to exposed positions in target proteins. Similarly, the intein component can also be fused to reporter proteins such as GFP, and to epitope or purification tags. In some embodiment, a paramyxovirus envelope protein is fused to a component of a universal adapter system (e.g., an intein). In further embodiments, the targeting moiety is fused to a component of a universal adapter system (e.g., an intein).

## B. Fusogen

### *1. Envelope attachment protein*

**[0106]** In some embodiments, the lipid particle includes a paramyxovirus envelope attachment protein exposed on the surface of the targeted lipid particle. In some embodiments, the paramyxovirus envelope attachment protein may be retargeted by fusion to one of the components of a universal adapter system, wherein a targeting moiety is fused to the other component of the universal adapter system (e.g. a targeting moiety that is an antibody or antigen-binding fragment, e.g. sdAb or scFv that binds to a target cell). In some embodiments, the paramyxovirus envelope attachment protein and an F protein provided herein together exhibit fusogenic activity to a target cell, such as to deliver an exogenous agent or nucleic acid exogenous agent to the target cell.

**[0107]** In some embodiments, the paramyxovirus envelope attachment protein is attached to a component universal adapter system (e.g., a tag polypeptide component or a catcher polypeptide component), and the targeting moiety is fused to the other of the components (e.g., a tag polypeptide component or a catcher polypeptide component).

**[0108]** The paramyxovirus envelope attachment protein can be linked directly or indirectly to a component universal adapter system. In particular embodiments, the component universal adapter is linked to the C-terminus (C-terminal amino acid) of the paramyxovirus envelope attachment protein or the biologically active portion thereof. In particular embodiments, the component universal adapter is linked to the N-terminus (N-terminal amino acid) of the paramyxovirus envelope attachment protein or the biologically active portion thereof. The linkage can be via a peptide linker, such as a flexible peptide linker.

**[0109]** In some embodiments, the C-terminus of the component universal adapter system is attached to the C-terminus of the paramyxovirus envelope attachment protein or biologically active portion thereof. In some embodiments, the N-terminus of the binding domain is exposed on the exterior surface

of the lipid bilayer. In some embodiments, the N-terminus of the component universal adapter system binds to a cognate component of the universal adapter system.

**[0110]** In some embodiments, the paramyxovirus envelope attachment protein or functionally active variant or biologically active portion thereof is linked directly to the bind component of the universal adapter system. In some embodiments, the paramyxovirus envelope attachment protein is a fusion protein that has the following structure: (N'- component of the universal adapter system -C')-(C'- paramyxovirus envelope attachment protein -N').

**[0111]** In some embodiments, the paramyxovirus envelope attachment protein is linked indirectly via a linker to the component of the universal adapter system. In some embodiments, the linker is a peptide linker. In some embodiments, the linker is a chemical linker.

**[0112]** In some embodiments, the linker is a peptide linker and the paramyxovirus envelope attachment protein is a fusion protein containing the paramyxovirus envelope attachment protein linked via a peptide linker to the component of the universal adapter system.

**[0113]** In some embodiments, the peptide linker is up to 65 amino acids in length. In some embodiments, the peptide linker comprises from or from about 2 to 65 amino acids, 2 to 60 amino acids, 2 to 56 amino acids, 2 to 52 amino acids, 2 to 48 amino acids, 2 to 44 amino acids, 2 to 40 amino acids, 2 to 36 amino acids, 2 to 32 amino acids, 2 to 28 amino acids, 2 to 24 amino acids, 2 to 20 amino acids, 2 to 18 amino acids, 2 to 14 amino acids, 2 to 12 amino acids, 2 to 10 amino acids, 2 to 8 amino acids, 2 to 6 amino acids, 6 to 65 amino acids, 6 to 60 amino acids, 6 to 56 amino acids, 6 to 52 amino acids, 6 to 48 amino acids, 6 to 44 amino acids, 6 to 40 amino acids, 6 to 36 amino acids, 6 to 32 amino acids, 6 to 28 amino acids, 6 to 24 amino acids, 6 to 20 amino acids, 6 to 18 amino acids, 6 to 14 amino acids, 6 to 12 amino acids, 6 to 10 amino acids, 6 to 8 amino acids, 8 to 65 amino acids, 8 to 60 amino acids, 8 to 56 amino acids, 8 to 52 amino acids, 8 to 48 amino acids, 8 to 44 amino acids, 8 to 40 amino acids, 8 to 36 amino acids, 8 to 32 amino acids, 8 to 28 amino acids, 8 to 24 amino acids, 8 to 20 amino acids, 8 to 18 amino acids, 8 to 14 amino acids, 8 to 12 amino acids, 8 to 10 amino acids, 10 to 65 amino acids, 10 to 60 amino acids, 10 to 56 amino acids, 10 to 52 amino acids, 10 to 48 amino acids, 10 to 44 amino acids, 10 to 40 amino acids, 10 to 36 amino acids, 10 to 32 amino acids, 10 to 28 amino acids, 10 to 24 amino acids, 10 to 20 amino acids, 10 to 18 amino acids, 10 to 14 amino acids, 10 to 12 amino acids, 12 to 65 amino acids, 12 to 60 amino acids, 12 to 56 amino acids, 12 to 52 amino acids, 12 to 48 amino acids, 12 to 44 amino acids, 12 to 40 amino acids, 12 to 36 amino acids, 12 to 32 amino acids, 12 to 28 amino acids, 12 to 24 amino acids, 12 to 20 amino acids, 12 to 18 amino acids, 12 to 14 amino acids, 14 to 65 amino acids, 14 to 60 amino acids, 14 to 56 amino acids, 14 to 52 amino acids, 14 to 48 amino acids, 14 to 44 amino acids, 14 to 40 amino acids, 14 to 36 amino acids, 14 to 32 amino acids, 14 to 28 amino acids, 14 to 24 amino acids, 14 to 20 amino acids, 14 to 18 amino acids, 18 to 65 amino acids, 18 to 60 amino acids, 18 to 56 amino acids, 18 to 52 amino acids, 18 to 48 amino acids, 18 to 44 amino acids, 18 to 40 amino acids, 18 to 36 amino acids, 18 to 32 amino acids, 18 to 28 amino acids, 18 to 24 amino

acids, 18 to 20 amino acids, 20 to 65 amino acids, 20 to 60 amino acids, 20 to 56 amino acids, 20 to 52 amino acids, 20 to 48 amino acids, 20 to 44 amino acids, 20 to 40 amino acids, 20 to 36 amino acids, 20 to 32 amino acids, 20 to 28 amino acids, 20 to 26 amino acids, 20 to 24 amino acids, 24 to 65 amino acids, 24 to 60 amino acids, 24 to 56 amino acids, 24 to 52 amino acids, 24 to 48 amino acids, 24 to 44 amino acids, 24 to 40 amino acids, 24 to 36 amino acids, 24 to 32 amino acids, 24 to 30 amino acids, 24 to 28 amino acids, 28 to 65 amino acids, 28 to 60 amino acids, 28 to 56 amino acids, 28 to 52 amino acids, 28 to 48 amino acids, 28 to 44 amino acids, 28 to 40 amino acids, 28 to 36 amino acids, 28 to 34 amino acids, 28 to 32 amino acids, 32 to 65 amino acids, 32 to 60 amino acids, 32 to 56 amino acids, 32 to 52 amino acids, 32 to 48 amino acids, 32 to 44 amino acids, 32 to 40 amino acids, 32 to 38 amino acids, 32 to 36 amino acids, 36 to 65 amino acids, 36 to 60 amino acids, 36 to 56 amino acids, 36 to 52 amino acids, 36 to 48 amino acids, 36 to 44 amino acids, 36 to 40 amino acids, 40 to 65 amino acids, 40 to 60 amino acids, 40 to 56 amino acids, 40 to 52 amino acids, 40 to 48 amino acids, 40 to 44 amino acids, 44 to 65 amino acids, 44 to 60 amino acids, 44 to 56 amino acids, 44 to 52 amino acids, 44 to 48 amino acids, 48 to 65 amino acids, 48 to 60 amino acids, 48 to 56 amino acids, 48 to 52 amino acids, 50 to 65 amino acids, 50 to 60 amino acids, 50 to 56 amino acids, 50 to 52 amino acids, 54 to 65 amino acids, 54 to 60 amino acids, 54 to 56 amino acids, 58 to 65 amino acids, 58 to 60 amino acids, or 60 to 65 amino acids. In some embodiments, the peptide linker is a polypeptide that is 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, or 65 amino acids in length.

**[0114]** In particular embodiments, the linker is a flexible peptide linker. In some such embodiments, the linker is 1-20 amino acids, such as 1-20 amino acids predominantly composed of glycine. In some embodiments, the linker is 1-20 amino acids, such as 1-20 amino acids predominantly composed of glycine and serine. In some embodiments, the linker is a flexible peptide linker containing amino acids Glycine and Serine, referred to as GS-linkers. In some embodiments, the peptide linker includes the sequences GS, GGS, GGGGS (SEQ ID NO:7), GGGGGS (SEQ ID NO:8) or combinations thereof. In some embodiments, the polypeptide linker has the sequence (GGS)n, wherein n is 1 to 10. In some embodiments, the polypeptide linker has the sequence (GGGGS)n, (SEQ ID NO:9) wherein n is 1 to 10. In some embodiments, the polypeptide linker has the sequence (GGGGGS)n (SEQ ID NO:10), wherein n is 1 to 6.

**[0115]** Provided herein are lipid particles comprising a paramyxovirus envelope attachment protein. Paramyxoviral attachment proteins are type II transmembrane glycoproteins that are designated as hemagglutinin-neuraminidase (HN), hemagglutinin (H), or glycoproteins (G), depending on two characteristics; the ability to agglutinate erythrocytes (hemagglutination) and the presence or absence of neuraminidase activity (cleavage of sialic acid). Specifically, the HN attachment glycoprotein is characteristic of the Avulavirus, Respirovirus, and Rubulavirus genera, the H attachment glycoproteins

are found in members of the Morbillivirus genus, while the G attachment glycoproteins are utilized by the viruses of the genus Henipavirus and the Pneumovirinae subfamily. The geometries of HN, H, or G glycoproteins possess high structural similarity, however although H and G glycoproteins are capable of recognizing protein receptors, they lack neuraminidase activity.

**[0116]** Paramyxoviral attachment glycoproteins contain a short N-terminal cytoplasmic tail, a transmembrane domain, and an extracellular domain containing an extracellular stalk and a globular head. The N-terminal cytoplasmic domain is within the inner lumen of the lipid bilayer and the C-terminal portion is the extracellular domain that is exposed on the outside of the lipid bilayer. The receptor binding and antigenic sites reside on the extracellular domain. Regions of the stalk in the C-terminal region have been shown to be involved in interactions with the F protein and triggering of fusion with a target cell membrane (Liu et al. 2015 J of Virology 89:1838). The F protein undergoes significant conformational change that facilitates the insertion of the fusion peptide into target membranes, bringing the two HR regions together in the formation of a six-helix bundle structure or trimer-of-hairpins during or immediately following fusion of virus and cell membranes (Bishop et al. 2008. J of Virology 82(22): 11398-11409). The cytoplasmic tails play a role in particle formation, incorporation into packaged particles, and serves as a signal peptide to modulate protein maturation and surface transport (Sawatsky et al. 2016. J of Virology 97:1066-1076).

a. G proteins

**[0117]** The attachment G proteins are type II transmembrane glycoproteins containing an N-terminal cytoplasmic tail (e.g. corresponding to amino acids 1-49 of SEQ ID NO:5), a transmembrane domain (e.g. corresponding to amino acids 50-70 of SEQ ID NO:5), and an extracellular domain containing an extracellular stalk (e.g. corresponding to amino acids 71-187 of SEQ ID NO:5), and a globular head (corresponding to amino acids 188-602 of SEQ ID NO:5). The N-terminal cytoplasmic domain is within the inner lumen of the lipid bilayer and the C-terminal portion is the extracellular domain that is exposed on the outside of the lipid bilayer. Regions of the stalk in the C-terminal region (e.g. corresponding to amino acids 71-187 of SEQ ID NO: 5) have been shown to be involved in interactions with F protein and triggering of F protein fusion (Liu et al. 2015 J of Virology 89:1838). In wild-type G protein, the globular head mediates receptor binding to henipavirus entry receptors Ephrin B2 and Ephrin B3, but is dispensable for membrane fusion (Brandel-Tretheway et al. Journal of Virology. 2019. 93(13)e00577-19). In some embodiments herein, tropism of the G protein is altered by linkage of the G protein or biologically active fragment thereof (e.g. cytoplasmic truncation) to a component of a universal adapter system. Binding of the G protein to a binding partner can trigger fusion mediated by a compatible F protein or biologically active portion thereof. G protein sequences disclosed herein are predominantly disclosed as expressed sequences including an N-terminal methionine required for start of translation. As such N-terminal methionines are commonly cleaved co- or post-translationally, the mature protein

sequences for all G protein sequences disclosed herein are also contemplated as lacking the N-terminal methionine.

**[0118]** G glycoproteins are highly conserved between henipavirus species. For example, the G protein of NiV and HeV viruses share 79% amino acids identity. Studies have shown a high degree of compatibility among G proteins with F proteins of different species as demonstrated by heterotypic fusion activation (Brandel-Tretheway et al. Journal of Virology. 2019. In particular embodiments, the F protein or the functionally active variant or biologically active portion thereof retains fusogenic activity in conjunction with a G protein as provided. Fusogenic activity includes the activity of the variant F protein in conjunction with a G protein to promote or facilitate fusion of two membrane lumens, such as the lumen of the lipid particle provided herein (e.g. having embedded in its lipid bilayer, such as exposed on its surface, a G protein and a variant F protein), and a cytoplasm of a target cell, e.g. a cell that contains a surface receptor or molecule that is recognized or bound by the G protein.

**[0119]** Exemplary Henipavirus protein G sequences are provided in **Table 2B**

**Table 2B**. Henipavirus protein G sequence clusters. Column 1, Genbank ID includes the Genbank ID of the whole genome sequence of the virus that is the centroid sequence of the cluster. Column 2, nucleotides of CDS provides the nucleotides corresponding to the CDS of the gene in the whole genome. Column 3, Full Gene Name, provides the full name of the gene including Genbank ID, virus species, strain, and protein name. Column 4, Sequence, provides the amino acid sequence of the gene. Column 5, #Sequences/Cluster, provides the number of sequences that cluster with this centroid sequence. Column 6 provides the SEQ ID numbers for the described sequences.

| **Table 2B: Exemplary Henipavirus protein G sequences** | | | | |
|---|---|---|---|---|
| **Genbank ID** | **Nucleotides of CDS** | **Full sequence ID** | **Sequence** | **SEQ ID NO** |
| AF017149 | 8913-10727 | gb:AF017149\|Organism:Hendravirus\|Strain Name:UNKNOWN-AF017149\|Protein Name:glycoprotein\|Gene Symbol:G | MMADSKLVSLNNNLSGKIKDQGKVIKNYYGT<br>MDIKKINDGLLDSKILGAFNTVIALLGSIIIIVM<br>NIMIIQNYTRTTDNQALIKESLQSVQQQIKALT<br>DKIGTEIGPKVSLIDTSSTITIPANIGLLGSKISQS<br>TSSINENVNDKCKFTLPPLKIHECNISCPNPLPF<br>REYRPISQGVSDLVGLPNQICLQKTTSTILKPR<br>LISYTLPINTREGVCITDPLLAVDNGFFAYSHL<br>EKIGSCTRGIAKQRIIGVGEVLDRGDKVPSMF<br>MTNVWTPPNPSTIHHCSSTYHEDFYYTLCAVS<br>HVGDPILNSTSWTESLSLIRLAVRPKSDSGDYN<br>QKYIAITKVERGKYDKVMPYGPSGIKQGDTLY<br>FPAVGFLPRTEFQYNDSNCPIIHCKYSKAENCR<br>LSMGVNSKSHYILRSGLLKYNLSLGGDIILQFI<br>EIADNRLTIGSPSKIYNSLGQPVFYQASYSWDT<br>MIKLGDVDTVDPLRVQWRNNSVISRPGQSQC<br>PRFNVCPEVCWEGTYNDAFLIDRLNWVSAGV<br>YLNSNQTAENPVFAVFKDNEILYQVPLAEDDT | 270 |

| | | | | |
|---|---|---|---|---|
| | | | NAQKTITDCFLLENVIWCISLVEIYDTGDSVIRP KLFAVKIPAQCSES | |
| AF2 1230 2 | 8943-10751 | gb:AF212302\|Organism:Nipah virus\|Strain Name:UNKNOWN-AF212302\|Protein Name:attachment glycoprotein\|Gene Symbol:G | MPAENKKVRFENTTSDKGKIPSKVIKSYYGTM DIKKINEGLLDSKILSAFNTVIALLGSIVIIVMNI MIIQNYTRSTDNQAVIKDALQGIQQQIKGLAD KIGTEIGPKVSLIDTSSTITIPANIGLLGSKISQST ASINENVNEKCKFTLPPLKIHECNISCPNPLPFR EYRPQTEGVSNLVGLPNNICLQKTSNQILKPKL ISYTLPVVGQSGTCITDPLLAMDEGYFAYSHL ERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFM TNVWTPPNPNTVYHCSAVYNNEFYYVLCAVS TVGDPILNSTYWSGSLMMTRLAVKPKSNGGG YNQHQLALRSIEKGRYDKVMPYGPSGIKQGD TLYFPAVGFLVRTEFKYNDSNCPITKCQYSKP ENCRLSMGIRPNSHYILRSGLLKYNLSDGENP KVVFIEISDQRLSIGSPSKIYDSLGQPVFYQASF SWDTMIKFGDVLTVNPLVVNWRNNTVISRPG QSQCPRFNTCPEICWEGVYNDAFLIDRINWISA GVFLDSNQTAENPVFTVFKDNEILYRAQLASE DTNAQKTITNCFLLKNKIWCISLVEIYDTGDN VIRPKLFAVKIPEQCT | 5 |
| JQ00 1776 | 8170-10275 | gb:JQ001776:8170-10275\|Organism:Cedar virus\|Strain Name:CG1a\|Protein Name:attachment glycoprotein\|Gene Symbol:G | MLSQLQKNYLDNSNQQGDKMNNPDKKLSVN FNPLELDKGQKDLNKSYYVKNKNYNVSNLLN ESLHDIKFCIYCIFSLLIIITIINIITISIVITRLKVHE ENNGMESPNLQSIQDSLSSLTNMINTEITPRIGI LVTATSVTLSSSINYVGTKTNQLVNELKDYIT KSCGFKVPELKLHECNISCADPKISKSAMYST NAYAELAGPPKIFCKSVSKDPDFRLKQIDYVIP VQQDRSICMNNPLLDISDGFFTYIHYEGINSCK KSDSFKVLLSHGEIVDRGDYRPSLYLLSSHYH PYSMQVINCVPVTCNQSSFVFCHISNNTKTLD NSDYSSDEYYITYFNGIDRPKTKKIPINNMTAD NRYIHFTFSGGGGVCLGEEFIIPVTTVINTDVFT HDYCESFNCSVQTGKSLKEICSESLRSPTNSSR YNLNGIMIISQNNMTDFKIQLNGITYNKLSFGS PGRLSKTLGQVLYYQSSMSWDTYLKAGFVEK WKPFTPNWMNNTVISRPNQGNCPRYHKCPEI CYGGTYNDIAPLDLGKDMYVSVILDSDQLAE NPEITVFNSTTILYKERVSKDELNTRSTTTSCFL FLDEPWCISVLETNRFNGKSIRPEIYSYKIPKYC | 271 |
| NC_ 0252 56 | 9117-11015 | gb:NC_025256:9117-11015\|Organism:Bat Paramyxovirus Eid_hel/GH-M74a/GHA/2009\|Strain Name:BatPV/Eid_hel/GH-M74a/GHA/2009\|Protein Name:glycoprot | MPQKTVEFINMNSPLERGVSTLSDKKTLNQSK ITKQGYFGLGSHSERNWKKQKNQNDHYMTV STMILEILVVLGIMFNLIVLTMVYYQNDNINQR MAELTSNITVLNLNLNQLTNKIQREIIPRITLID TATTITIPSAITYILATLTTRISELLPSINQKCEFK TPTLVLNDCRINCTPPLNPSDGVKMSSLATNL VAHGPSPCRNFSSVPTIYYYRIPGLYNRTALDE RCILNPRLTISSTKFAYVHSEYDKNCTRGFKYY ELMTFGEILEGPEKEPRMFSRSFYSPTNAVNY HSCTPIVTVNEGYFLCLECTSSDPLYKANLSNS TFHLVILRHNKDEKIVSMPSFNLSTDQEYVQII PAEGGGTAESGNLYFPCIGRLLHKRVTHPLCK KSNCSRTDDESCLKSYYNQGSPQHQVVNCLIR | 272 |

| | | | | |
|---|---|---|---|---|
| | | ein\|Gene Symbol:G | IRNAQRDNPTWDVITVDLTNTYPGSRSRIFGSF SKPMLYQSSVSWHTLLQVAEITDLDKYQLDW LDTPYISRPGGSECPFGNYCPTVCWEGTYNDV YSLTPNNDLFVTVYLKSEQVAENPYFAIFSRD QILKEFPLDAWISSARTTTISCFMFNNEIWCIAA LEITRLNDDIIRPIYYSFWLPTDCRTPYPHTGK MTRVPLRSTYNY | |
| NC_ 025352 | 8716-11257 | gb:NC_025352: 8716-11257\|Organism:Mojiang virus\|Strain Name:Tonggua n1\|Protein Name:attachment glycoprotein\|Gene Symbol:G | MATNRDNTITSAEVSQEDKVKKYYGVETAEK VADSISGNKVFILMNTLLILTGAIITITLNITNLT AAKSQQNMLKIIQDDVNAKLEMFVNLDQLVK GEIKPKVSLINTAVSVSIPGQISNLQTKFLQKY VYLEESITKQCTCNPLSGIFPTSGPTYPPTDKPD DDTTDDDKVDTTIKPIEYPKPDGCNRTGDHFT MEPGANFYTVPNLGPASSNSDECYTNPSFSIGS SIYMFSQEIRKTDCTAGEILSIQIVLGRIVDKGQ QGPQASPLLVWAVPNPKIINSCAVAAGDEMG WVLCSVTLTAASGEPIPHMFDGFWLYKLEPDT EVVSYRITGYAYLLDKQYDSVFIGKGGGIQKG NDLYFQMYGLSRNRQSFKALCEHGSCLGTGG GGYQVLCDRAVMSFGSEESLITNAYLKVNDL ASGKPVIIGQTFPPSDSYKGSNGRMYTIGDKY GLYLAPSSWNRYLRFGITPDISVRSTTWLKSQ DPIMKILSTCTNTDRDMCPEICNTRGYQDIFPL SEDSEYYTYIGITPNNGGTKNFVAVRDSDGHI ASIDILQNYYSITSATISCFMYKDEIWCIAITEG KKQKDNPQRIYAHSYKIRQMCYNMKSATVTV GNAKNITIRRY | 273 |

**[0120]** In some embodiments, the G protein has a sequence set forth in any of SEQ ID NOS: 5, 270, 271, 272, or 273 or is a functionally active variant or biologically active portion thereof that has a sequence that is at least at or about 80%, at least at or about 81%, at least at or about 82%, at least at or about 83%, at least at or about 84%, at least at or about 85%, at least at or about 86%, at least at or about 87%, at least at or about 88%, at least at or about 89%, at least at or about 90%, at least at or about 91%, at least at or about 92%, at least at or about 93%, at least at or about 94%, at least at or about 95%, at least at or about 96%, at least at or about 97%, at least at or about 98%, or at least at or about 99% identical to any one of SEQ ID NOS: 5, 270, 271, 272, or 273.

**[0121]** In particular embodiments, the G protein or functionally active variant or biologically active portion is a protein that retains fusogenic activity in conjunction with a paramyxovirus F protein, such as a NiV-F protein described herein. Fusogenic activity includes the activity of the G protein in conjunction with a F protein to promote or facilitate fusion of two membrane lumens, such as the lumen of the targeted lipid particle having embedded in its lipid bilayer a F and G protein, and a cytoplasm of a target cell, e.g. a cell that contains a surface receptor or molecule that is recognized or bound by the targeted envelope protein via fusion to a component of a universal adapter system. In some embodiments, the F protein and G protein are from the same species (e.g. NiV-G and NiV-F).

**[0122]** In some embodiments, the NiV-G is a variant NiV-G proteins that contain an altered cytoplasmic tail compared to native NiV-G (e.g. SEQ ID NO:49) that are or can be incorporated into a lipid particle, such as a viral particle, including a lentiviral particle or lentiviral-like particle. The cytoplasmic tail of NiV-G corresponds to amino acids 1-45 of SEQ ID NO:49. In some cases, it is understood that the N-terminal methionine of NiV-G, or a variant NiV-G, as described herein can be cleaved and the cytoplasmic tail lacks an initial N-terminal methionine. For instance, in some embodiments, the cytoplasmic tail of wild-type NiV-G may correspond to amino acids 2-45 of SEQ ID NO:49, and the variant NiV-G protein contains a cytoplasmic tail that is altered compared to amino acids 2-45 of SEQ ID NO:49. In some embodiments, the variant NiV-G contains a modified cytoplasmic tail in which the native cytoplasmic tail is truncated or is replaced by a heterologous cytoplasmic tail. Exemplary alternative cytoplasmic tails are set forth in SEQ ID NOs 84-269. In some embodiments, the variant NiV-G contains a modified cytoplasmic tail in which the native cytoplasmic is replaced by a heterologous cytoplasmic tail set forth in any one of SEQ ID NOs 84-269.

**[0123]** Non-limiting examples of variant NiV-G proteins, including truncated NiV-G or NiV-G with a altered or modified cytoplasmic tail, are described in WO2013148327, WO2017182585, or PCT/US2022/081872. Further exemplary variant NiV-G proteins are described in Bender et al. 2016 PLoS Pathol 12(6):e1005641.

**[0124]** In some embodiments, the G protein is a variant G protein that is a functionally active variant or biologically active portion containing one or more amino acid mutations, such as one or more amino acid insertions, deletions, substitutions or truncations. In some embodiments, the mutations described herein relate to amino acid insertions, deletions, substitutions or truncations of amino acids compared to a reference G protein sequence. In some embodiments, the reference G protein sequence is the wild-type sequence of a G protein or a biologically active portion thereof. In some embodiments, at least one functionally active variant or the biologically active portion thereof is a variant of a wild-type Hendra (HeV) virus G protein, a wild-type Nipah (NiV) virus G-protein (NiV-G), a wild-type Cedar (CedPV) virus G-protein, a wild-type Mojiang virus G-protein, a wild-type bat Paramyxovirus G-protein or biologically active portion thereof. In some embodiments, the wild-type G protein has the sequence set forth in any one of SEQ ID NOS: 5, 270, 271, 272, or 273.

**[0125]** In some embodiments, the G protein is a variant G protein that is a biologically active portion that is an N-terminally and/or C-terminally truncated fragment of a wild-type Hendra (HeV) virus G protein, a wild-type Nipah (NiV) virus G-protein (NiV-G), a wild-type Cedar (CedPV) virus G-protein, a wild-type Mojiang virus G-protein, a wild-type bat Paramyxovirus G-protein. In particular embodiments, the truncation is an N-terminal truncation of all or a portion of the cytoplasmic domain. In some embodiments, at least one variant G protein is a biologically active portion that is truncated and lacks up to 49 contiguous amino acid residues at or near the N-terminus of the wild-type G protein, such as a wild-type G protein set forth in any one of SEQ ID NOS: 5, 270, 271, 272, or 273. In some

embodiments, at least one variant G protein is truncated and lacks up to 49 contiguous amino acids, such as up to 49, 48, 47, 46, 45, 44, 43, 42, 41, 40, 30, 38, 37, 36, 35, 34, 33, 32, 31, 30, 29, 28, 27, 26, 25, 24, 23, 22, 21, 20, 19, 18, 17, 16, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2 or 1 contiguous amino acids at the N-terminus of the wild-type G protein.

**[0126]** In some embodiments, the G protein is a wild-type Nipah virus G (NiV-G) protein, or is a functionally active variant or biologically active portion thereof. In some embodiments, the G protein is a NiV-G protein that has the sequence set forth in SEQ ID NO:5, or is a functional variant or a biologically active portion thereof that has an amino acid sequence having at least at or about 80%, at least at or about 81%, at least at or about 82%, at least at or about 83%, at least at or about 84%, at least at or about 85%, at least at or about 86%, at least at or about 87%, at least at or about 88%, at least at or about 89%, at least at or about 90%, at least at or about 91%, at least at or about 92%, at least at or about 93%, at least at or about 94%, at least at or about 95%, at or about 96%, at least at or about 97%, at least at or about 98%, at least at or about 99% sequence identity to SEQ ID NO:5.

**[0127]** In some embodiments, the variant NiV-G comprises a modified cytoplasmic tail which comprises a truncated cytoplasmic tail from a glycoprotein from the same Nipah virus. In some embodiments, the variant NiV-G contains a modified cytoplasmic tail in which at least a part of the native cytoplasmic tail (e.g. corresponding to amino acids 1-45 of SEQ ID NO:49) is a truncated portion thereof from a glycoprotein from Nipah Virus. In some embodiments, the cytoplasmic tail is a truncated portion thereof that is at least 5 amino acids in length. from or from about 5-44, from or from about 5-40, from or from about 5-30, from or from about 5-20, from or from about 5-10, from or from about 10-44, from or from about 10-40, from or from about 10-30, from or from about 10-20, from or from about 20-44, from or from about 20-40, from or from about 20-30, from or from about 30-44, from or from about 30-40, from or from about 40-44amino acids in length. In some embodiments, the truncated portion thereof is 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43 or 44 amino acids in length. In some embodiments, the variant NiV-G has a cytoplasmic tail that is a truncated NiV-G cytoplasmic tail. In some embodiments, the truncated NiV-G cytoplasmic tail has a deletion of up to 40, up to 35, up to 30, up to 29, up to 28, up to 27, up to 26, up to 25, up to 24, up to 23, up to 22, up to 21, up to 20, up to 19, up to 18, up to 17, up to 16, up to 15, or up to 14 contiguous amino acid residues at or near the N-terminus of the wild-type NiV-G cytoplasmic tail set forth in SEQ ID NO: 72. In some embodiments, the variant NiV-G has a deletion of between 5 and 41 contiguous amino acid residues at or near the N-terminus of the wild-type NiV-G protein cytoplasmic tail set forth in SEQ ID NO: 48. In some embodiments, the variant NiV-G has a deletion of between 26 and 40 contiguous amino acid residues at or near the N-terminus of the wild-type NiV-G protein cytoplasmic tail set forth in SEQ ID NO: 48.

**[0128]** In some embodiments, the G protein is a NiV-G protein that is a biologically active portion of a wild-type NiV-G. In some embodiments, the biologically active portion is an N-terminally truncated

fragment. In some embodiments, the NiV-G protein is truncated and lacks up to 5 contiguous amino acid residues at or near the N-terminus of the wild-type NiV-G protein, such as compared to wild-type NiV-G set forth in SEQ ID NO: 5. In some embodiments, the NiV-G protein is truncated and lacks up to 10 contiguous amino acid residues at or near the N-terminus of the wild-type NiV-G protein, such as compared to wild-type NiV-G set forth in SEQ ID NO: 5. In some embodiments, the NiV-G protein is truncated and lacks up to 15 contiguous amino acid residues at or near the N-terminus of the wild-type NiV-G protein, such as compared to wild-type NiV-G set forth in SEQ ID NO: 5. n some embodiments, the NiV-G protein is truncated and lacks up to 20 contiguous amino acid residues at or near the N-terminus of the wild-type NiV-G protein, such as compared to wild-type NiV-G set forth in SEQ ID NO: 5. In some embodiments, the NiV-G protein is truncated and lacks up to 25 contiguous amino acid residues at or near the N-terminus of the wild-type NiV-G protein, such as compared to wild-type NiV-G set forth in SEQ ID NO: 5. In some embodiments, the NiV-G protein is truncated and lacks up to 30 contiguous amino acid residues at or near the N-terminus of the wild-type NiV-G protein, such as compared to wild-type NiV-G set forth in SEQ ID NO: 5. In some embodiments, the NiV-G protein is truncated and lacks up to 35 contiguous amino acid residues at or near the N-terminus of the wild-type NiV-G protein, such as compared to wild-type NiV-G set forth in SEQ ID NO: 5. In some embodiments, the NiV-G protein (also called variant NiV-G) contains an N-terminal methionine.

**[0129]** In some embodiments, the variant NiV-G has a cytoplasmic tail deletion of amino acid residues 2-41, 2-40, 2-39, 2-38, 2-37, 2-36, 2-35, 2-34, 2-33, 2-32, 2-31, 2-30, 2-29, 2-28, 2-27, 2-26, 2-25, 2-22, 2-21, 2-16, 2-11, or 2-5 of SEQ ID NO:48. In some embodiments, the cytoplasmic tail is a truncated portion of the Nipah virus cytoplasmic tail set forth in any one of SEQ ID NOS: 50-72. In some embodiments, the cytoplasmic tail is a truncated portion of the Nipah virus cytoplasmic tail set forth in any one of SEQ ID NOS: 50-72 that lacks the N-terminal methionine. In some embodiments, the variant NiV-G has a sequence in which the cytoplasmic tail, such as set forth in any one of SEQ ID NOS: 6-28, is directly linked to the N-terminus of the sequence set forth in SEQ ID NO: 46. In some embodiments, the variant NiV-G has a sequence in which the cytoplasmic tail set forth in any one of SEQ ID NOS: 6-28 is directly linked to the N-terminus of the sequence set forth in SEQ ID NO: 47. In some embodiments, the truncated NiV-G cytoplasmic tail comprises is set forth in SEQ ID NO: 51, 57 or 63.

**[0130]** In some embodiments, the variant NiV-G comprises the sequence of amino acids set forth in SEQ ID NO: 51, 57, 63, 255, 264 or 265, or a sequence of amino acids that exhibits at least 85% sequence identity, at least 86% sequence identity, at least 87% sequence identity, at least 88% sequence identity, at least 89% sequence identity, at least 90% sequence identity, at least 91% sequence identity, at least 92% sequence identity, at least 93% sequence identity, at least 94% sequence identity, at least 95% sequence identity, at least 96% sequence identity, at least 97% sequence identity, at least 98% sequence identity, or at least 99% sequence identity to any one of SEQ ID NOs: 51, 57, 63, 255, 264 or 265. In

some embodiments, the variant NiV-G is the sequence of amino acids set forth in SEQ ID NO: 51, 57, 63, 255, 264 or 265.

**[0131]** In some embodiments, the variant NiV-G comprises a modified cytoplasmic tail which comprises a heterologous cytoplasmic tail or a truncated portion thereof from a glycoprotein from another virus. In some embodiments, the other virus is a member of the Kingdom Orthornavirae. In some embodiments, the other virus is a member of the family Paramyxoviridae, Rhabdoviridae, Arenaviridae, or Retroviridae. In some embodiments, the other virus is a member of the family Paramyxoviridae.

**[0132]** In some embodiments, the variant NiV-G contains a modified cytoplasmic tail in which at least a part of the native cytoplasmic tail (e.g. corresponding to amino acids 1-45 of SEQ ID NO:49) is replaced by a heterologous cytoplasmic tail or a truncated portion thereof from a glycoprotein from another virus from another virus or viral-associated protein. In some embodiments, the replaced cytoplasmic tail is a heterologous cytoplasmic tail or a truncated portion thereof that is at least 5 amino acids in length. In some embodiments, the replaced heterologous cytoplasmic tail or a truncated portion thereof is from or from about 5-180 amino acids in length, such as from or from about 5-150, from or from about 5-100, from or from about 5-75, from or from about 5-50, from or from about 5-40, from or from about 5-30, from or from about 5-20, from or from about 5-10, from or from about 10-150, from or from about 10-100, from or from about 10-75, from or from about 10-50, from or from about 10-40, from or from about 10-30, from or from about 10-20, from or from about 20-150, from or from about 20-100, from or from about 20-75, from or from about 20-50, from or from about 20-40, from or from about 20-30, from or from about 30-150, from or from about 30-100, from or from about 30-75, from or from about 30-50, from or from about 30-40, from or from about 40-150, from or from about 40-100, from or from about 40-75, from or from about 40-50, from or from about 50-150, from or from about 50-100, from or from about 50-75, from or from about 75-150, from or from about 75-100 or from or from about 100-150 amino acids in length. In some embodiments, the replaced heterologous cytoplasmic tail or a truncated portion thereof is 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39 or 40 amino acids in length. In some embodiments, the heterologous cytoplasmic tail or the truncated portion thereof is directly linked to the N-terminus of the sequence set forth in SEQ ID NO: 46.

**[0133]** In some embodiments, the heterologous cytoplasmic tail is a cytoplasmic tail or a truncated portion thereof from a glycoprotein from another virus, such as a paramyxovirus, a retrovirus, a filovirus, a rhabdovirus or an arenavirus. In some embodiments, the virus is a paramyxovirus other than a Nipah virus. For instance, the virus is a measles virus, Bat paramyxovirus, Cedar Virus, Canine Distemper Virus, Sendai virus, Hendra virus, Human Parainfluenza virus, or Newcastle Disease virus. In some embodiments, the replaced heterologous cytoplasmic tail is the native cytoplasmic tail or a truncated portion of the native cytoplasmic tail of another virus, such as a truncated portion of the cytoplasmic tail set forth in any one of SEQ ID NOS: 84-210. In some embodiments, the variant NiV-G has a sequence

in which the heterologous cytoplasmic tail or the truncated portion thereof set forth in any one of SEQ ID NOS: 84-210 is directly linked to the N-terminus of the sequence set forth in SEQ ID NO: 46. In some embodiments, the variant NiV-G contains mutations in the extracellular domain that reduce or abrogate binding to an Ephrin B2 or B3 corresponding to one or more of E501A, W504A, Q530A and E533A, with numbering of residues as set forth SEQ ID NO:5. In some embodiments, the variant NiV-G has a sequence in which the heterologous cytoplasmic tail or the truncated portion thereof set forth in any one of SEQ ID NOS: 84-210 is directly linked to the N-terminus of the sequence set forth in SEQ ID NO: 47. In some embodiments, it is understood that the heterologous cytoplasmic tail or the truncated portion thereof may include any sequence set forth in any one of SEQ ID NOS: 84-210 that lacks the N-terminal methionine.

**[0134]** In some embodiments, the virus is a retrovirus. For instance, the virus may be a baboon endogenous virus (BaEV), Gibbon Ape Leukemia virus (GaLV), murine leukemia virus, or human immunodeficiency virus 1 (HIV-1). In some embodiments, the replaced heterologous cytoplasmic tail is the native cytoplasmic tail or a truncated portion of the native cytoplasmic tail of another virus, such as set forth in any one of SEQ ID NOS: 211-212, 218-221, 223-226, or 229-243. In some embodiments, the variant NiV-G has a sequence in which the heterologous cytoplasmic tail or the truncated portion thereof set forth in any one of SEQ ID NOS: 211-212, 218-221, 223-226, or 229-243 is directly linked to the N-terminus of the sequence set forth in SEQ ID NO: 46. In some embodiments, the variant NiV-G contains mutations in the extracellular domain that reduce or abrogate binding to an Ephrin B2 or B3 corresponding to one or more of E501A, W504A, Q530A and E533A, with numbering of residues as set forth SEQ ID NO:5. In some embodiments, the variant NiV-G has a sequence in which the heterologous cytoplasmic tail or the truncated portion thereof set forth in any one of SEQ ID NOS: 211-212, 218-221, 223-226, or 229-243 is directly linked to the N-terminus of the sequence set forth in SEQ ID NO: 47. In some embodiments, it is understood that the heterologous cytoplasmic tail or the truncated portion thereof may include any sequence set forth in any one of SEQ ID NOS: 211-212, 218-221, 223-226, or 229-243 that lacks the N-terminal methionine.

**[0135]** In some embodiments, the virus is a filovirus. For instance, the virus may be an Ebola virus (EboV). In some embodiments, the replaced heterologous cytoplasmic tail is the native cytoplasmic tail or a truncated portion of the native cytoplasmic tail of another virus, such as set forth in any one of SEQ ID NOS: 216 or 217. In some embodiments, the variant NiV-G has a sequence in which the heterologous cytoplasmic tail or the truncated portion thereof set forth in any one of SEQ ID NOS: 216 or 217 is directly linked to the N-terminus of the sequence set forth in SEQ ID NO: 46. In some embodiments, the variant NiV-G contains mutations in the extracellular domain that reduce or abrogate binding to an Ephrin B2 or B3 corresponding to one or more of E501A, W504A, Q530A and E533A, with numbering of residues as set forth SEQ ID NO:5. In some embodiments, the variant NiV-G has a sequence in which the heterologous cytoplasmic tail or the truncated portion thereof set forth in any one of SEQ ID NOS:

216 or 217 is directly linked to the N-terminus of the sequence set forth in SEQ ID NO: 47. In some embodiments, it is understood that the heterologous cytoplasmic tail or the truncated portion thereof may include any sequence set forth in any one of SEQ ID NOS: 216 or 217 that lacks the N-terminal methionine.

**[0136]** In some embodiments, the virus is a rhabdovirus. For instance, the virus may be Cocal vesiculovirus (Cocal) or vesicular stomatitis virus (VSV). In some embodiments, the replaced heterologous cytoplasmic tail is the native cytoplasmic tail or a truncated portion of the native cytoplasmic tail of another virus, such as set forth in any one of SEQ ID NOS: 214, 215, 227, or 228. In some embodiments, the variant NiV-G has a sequence in which the heterologous cytoplasmic tail or the truncated portion thereof set forth in any one of SEQ ID NOS: 114, 215, 227, or 228 is directly linked to the N-terminus of the sequence set forth in SEQ ID NO: 46. In some embodiments, the variant NiV-G contains mutations in the extracellular domain that reduce or abrogate binding to an Ephrin B2 or B3 corresponding to one or more of E501A, W504A, Q530A and E533A, with numbering of residues as set forth SEQ ID NO:5. In some embodiments, the variant NiV-G has a sequence in which the heterologous cytoplasmic tail or the truncated portion thereof set forth in any one of SEQ ID NOS: 214, 215, 227, or 228 is directly linked to the N-terminus of the sequence set forth in SEQ ID NO: 47. In some embodiments, it is understood that the heterologous cytoplasmic tail or the truncated portion thereof may include any sequence set forth in any one of SEQ ID NOS: 214, 215, 227, or 228 that lacks the N-terminal methionine.

**[0137]** In some embodiments, the virus is an arenavirus. For instance, the virus may be Lymphocytic choriomeningitis virus (LCMV). In some embodiments, the replaced heterologous cytoplasmic tail is the native cytoplasmic tail or a truncated portion of the native cytoplasmic tail of another virus, such as set forth in SEQ ID NOS: 222. In some embodiments, the variant NiV-G has a sequence in which the heterologous cytoplasmic tail or the truncated portion thereof set forth in SEQ ID NOS: 222 is directly linked to the N-terminus of the sequence set forth in SEQ ID NO: 46. In some embodiments, the variant NiV-G contains mutations in the extracellular domain that reduce or abrogate binding to an Ephrin B2 or B3 corresponding to one or more of E501A, W504A, Q530A and E533A, with numbering of residues as set forth SEQ ID NO:5. In some embodiments, the variant NiV-G has a sequence in which the heterologous cytoplasmic tail or the truncated portion thereof set forth in SEQ ID NOS: 178 is directly linked to the N-terminus of the sequence set forth in SEQ ID NO: 47. In some embodiments, it is understood that the heterologous cytoplasmic tail or the truncated portion thereof may include any sequence set forth in any one of SEQ ID NOS: 222 that lacks the N-terminal methionine.

**[0138]** In some embodiments, the NiV-G protein is truncated and lacks up to amino acid 34 at or near the N-terminus of the wild-type NiV-G protein, such as compared to wild-type NiV-G set forth in SEQ ID NO: 5. In some embodiments, the NiV-G protein contains an N-terminal methionine. In some

embodiments, the NiV-G protein lacks amino acids 2-34 as compared to wild-type NiV-G set forth in SEQ ID NO:5. In some embodiments, the NiV-G has the sequence set forth in SEQ ID NO:6.

**[0139]** In some embodiments, the variant NiV-G contains a heterologous cytoplasmic tail that is a cytoplasmic tail or a truncated portion thereof from a glycoprotein from CD63. In some embodiments, the heterologous cytoplasmic tail replaces at least a part of the native cytoplasmic tail of NiV-G (e.g. corresponding to amino acids 1-45 of SEQ ID NO:49). In some embodiments, the heterologous tail is a contiguous sequence of 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, or 25 N-terminal amino acids of the native cytoplasmic tail of CD63. In some embodiments, the native cytoplasmic tail of CD63 is set forth in SEQ ID NOs: 244, 245, or 246. In some embodiments, the heterologous cytoplasmic tail is a truncated portion of the CD63 cytoplasmic tail set forth in any one of SEQ ID NOS: 244-249. In some embodiments, the variant NiV-G has a sequence in which the heterologous cytoplasmic tail set forth in any one of SEQ ID NOS: 244-249 is directly linked to the N-terminus of the sequence set forth in SEQ ID NO: 46. In some embodiments, the variant NiV-G has a sequence in which the heterologous cytoplasmic tail set forth in any one of SEQ ID NOS: 244-249 is directly linked to the N-terminus of the sequence set forth in SEQ ID NO: 47.

**[0140]** In some embodiments, the variant NiV-G comprises a modified cytoplasmic tail which comprises a mutated cytoplasmic tail from a glycoprotein from the same Nipah virus. In some embodiments, the variant NiV-G contains a modified cytoplasmic tail in which at least a part of the native cytoplasmic tail (e.g. corresponding to amino acids 1-45 of SEQ ID NO:49) is a mutated portion thereof from a glycoprotein from Nipah Virus. In some embodiments, the cytoplasmic tail is a mutated portion of the Nipah virus cytoplasmic tail set forth in any one of SEQ ID NOS: 73-82. In some embodiments, it is understood that the truncated NiV-G cytoplasmic tail may include the sequence set forth in any one of SEQ ID NOS: 73-82 that lacks the N-terminal methionine. In some embodiments, the variant NiV-G has a sequence in which the cytoplasmic tail set forth in any one of SEQ ID NOS: 73-82 is directly linked to the N-terminus of the sequence set forth in SEQ ID NO: 46. In some embodiments, the variant NiV-G has a sequence in which the cytoplasmic tail set forth in any one of SEQ ID NOS: 73-82 is directly linked to the N-terminus of the sequence set forth in SEQ ID NO: 47.

**[0141]** In some embodiments, the G protein or the functionally active variant or biologically active portion thereof binds to Ephrin B2 or Ephrin B3. In some embodiments, the G protein is a mutant G protein, such as a truncated G protein as described and retains binding to Ephrin B2 or B3. Reference to retaining binding to Ephrin B2 or B3 includes binding that is similar to the level or degree of binding of the corresponding wild-type G protein, such as set forth in SEQ ID NO: 5, 270, 271, 272, or 273 such as at least 50%, at least 60%, at least 70%, at least 80% or at least 90% of the binding of the wild-type G protein.

**[0142]** In some embodiments, also provided are any of the provided variant NiV-G proteins that are re-targeted compared to the native tropism of NiV-G. For instance, mutations in NiV-G that completely

abrogate ephrinB2 and B3 binding, but that do not impact the association of this NiV-G with NiV-F, have been identified (Aguilar, et al. J Biol Chem. 2009;284(3):1628-1635.; Weise et al. J Virol. 2010;84(15):7634-764; Negrete et al.. J Virol. 2007;81(19):10804-10814; Negrete et al. PLoS Pathog. 2006; Guillaume et al., J. Virol 2006, 80 (15) 7546-7554). Thus, in provided aspects, a variant NiV-G protein provided herein may further contain a mutation in its extracellular domain to reduce or abrogate binding to Ephrin B2 and/B3. In some embodiments, the mutations can include one or more of mutations E501A, W504A, Q530A and E533A, with reference to numbering of wild-type NiV-G set forth in SEQ ID NO:49. In some embodiments, any of the provided variant NiV-G proteins may also be linked or fused to a binding molecule for targeted binding to a target molecule of interest. In some embodiments, the variant G protein is a fusion of a binding molecule with variant NiV-G, including a NiV-G with mutations to abrogate Ephrin B2 and/or Ephrin B3 binding. This could allow for altered G protein tropism allowing for targeting of other desired cell types that are not ephrinB2+ through the addition of the binding molecule directed against a different cell surface molecule.

**[0143]** In some embodiments, the G protein or the biologically thereof is a G protein that exhibits reduced binding for the native binding partner of a wild-type G protein. In some embodiments, the G protein or the biologically active portion thereof is a mutant of wild-type Niv-G (e.g., variant Niv-G) and exhibits reduced binding to one or both of the native binding partners Ephrin B2 or Ephrin B3. In some embodiments, the G-protein or the biologically active portion, such as a variant NiV-G protein, exhibits reduced binding to the native binding partner. In some embodiments, the reduced binding to Ephrin B2 or Ephrin B3 is reduced by greater than at or about 5%, at or about 10%, at or about 15%, at or about 20%, at or about 25%, at or about 30%, at or about 40%, at or about 50%, at or about 60%, at or about 70%, at or about 80%, at or about 90%, or at or about 100%.

**[0144]** In some embodiments, the mutations can improve transduction efficiency. In some embodiments, the mutations allow for specific targeting of other desired cell types that are not Ephrin B2 or Ephrin B3. In some embodiments, the mutations result in at least the partial inability to bind at least one natural receptor, such has reduced the binding to at least one of Ephrin B2 or Ephrin B3. In some embodiments, the mutations described herein interfere with natural receptor recognition.

**[0145]** In some embodiments, the G protein contains one or more amino acid substitutions in a residue that is involved in the interaction with one or both of Ephrin B2 and Ephrin B3. In some embodiments, the amino acid substitutions correspond to mutations E501A, W504A, Q530A and E533A with reference to numbering set forth in SEQ ID NO:5. In some embodiments, the G protein is a mutant G protein containing one or more amino acid substitutions selected from the group consisting of E501A, W504A, Q530A and E533A with reference to numbering set forth in SEQ ID NO:5. In some embodiments, the G protein is a mutant G protein that contains one or more amino acid substitutions elected from the group consisting of E501A, W504A, Q530A and E533A with reference to SEQ ID NO:5 and is a biologically active portion thereof containing an N-terminal truncation.

**[0146]** In particular embodiments, at least one G protein has the sequence of amino acids set forth in any one of SEQ ID NOs: 46-273, or is a functionally active variant thereof or a biologically active portion thereof that retains binding and/or fusogenic activity. In some embodiments, the functionally active variant comprises an amino acid sequence having at least at or about 80%, at least at or about 85%, at least at or about 90%, at least at or about 91%, at least at or about 92%, at least at or about 93%, at least at or about 94%, at least at or about 95%, at or about 96%, at least at or about 97%, at least at or about 98%, or at least at or about 99% sequence identity to any one of SEQ ID NOs: 46-273 and retains fusogenic activity in conjunction with a variant NiV-F protein as described. In some embodiments, at least one G protein is a variant G protein that comprises the amino acid sequence of any one of SEQ ID NOs: 46-273.

**[0147]** In some embodiments, any of the provided lipid particles (lentiviral vectors) may also contain an F protein, such as a NiV-F protein, such as a full-length NiV-F protein or a biologically active portion thereof or a variant thereof. For instance, also provided herein are viral particles or viral-like particles, such as lentiviral particles or lentiviral-like particles, that are pseudotyped with any of the provided variant NiV-G proteins and a NiV-F protein, such as a full-length NiV-F protein or a biologically active portion or a variant thereof. Exemplary NiV-F proteins are further described in Section I.B.2.

**[0148]** In some embodiments, the G protein or functionally active variant or biologically active portion thereof, is a protein that retains fusogenic activity in conjunction with other retargeted attachment proteins, such as more than one G protein expressed as a multimer on the lipid bi-layer. In particular embodiments, the paramyxovirus attachment protein, such as at least one G protein or functionally active variant or biologically active portion thereof, is a protein that retains fusogenic activity in conjunction with other retargeted attachment proteins, such as more than one G protein expressed as a multimer on the lipid bi-layer. Fusogenic activity includes the activity of the paramyxovirus envelope attachment protein in conjunction with a protein that is a paramyxovirus fusion protein (e.g., an F protein) to promote or facilitate fusion of two membrane lumens, such as the lumen of the targeted lipid particle having embedded in its lipid bilayer at a paramyxovirus envelope attachment protein and paramyxovirus fusion protein (e.g., F and G proteins), and a cytoplasm of a target cell, e.g. a cell that contains a surface receptor or molecule that is recognized or bound by the targeted envelope protein.

**[0149]** In some embodiments, also provided are any of the provided variant NiV-G proteins that are re-targeted compared to the native tropism of NiV-G. For instance, mutations in NiV-G that completely abrogate ephrinB2 and B3 binding, but that do not impact the association of this NiV-G with NiV-F, have been identified (Aguilar, et al. J Biol Chem. 2009;284(3):1628-1635.; Weise et al. J Virol. 2010;84(15):7634-764; Negrete et al.. J Virol. 2007;81(19):10804-10814; Negrete et al. PLoS Pathog. 2006; Guillaume et al., J. Virol 2006, 80 (15) 7546-7554). Thus, in provided aspects, a variant NiV-G protein provided herein may further contain a mutation in its extracellular domain to reduce or abrogate

binding to Ephrin B2 and/B3. In some embodiments, the mutations can include one or more of mutations E501A, W504A, Q530A and E533A, with reference to numbering of wild-type NiV-G set forth in SEQ ID NO:5. In some embodiments, any of the provided variant NiV-G proteins may also be linked or fused to a binding molecule for targeted binding to a target molecule of interest.

**[0150]** In some embodiments, the variant G protein is a fusion of a binding molecule with variant NiV-G, including a NiV-G with mutations to abrogate Ephrin B2 and/or Ephrin B3 binding. This could allow for altered G protein tropism allowing for targeting of other desired cell types that are not ephrinB2+ through the addition of the binding molecule directed against a different cell surface molecule.

**[0151]** In particular embodiments, the G protein has the sequence of amino acids set forth in SEQ ID NO: 6, or is a functionally active variant thereof or a biologically active portion thereof that retains binding and/or fusogenic activity. In some embodiments, the functionally active variant comprises an amino acid sequence having at least at or about 80%, at least at or about 85%, at least at or about 90%, at least at or about 91%, at least at or about 92%, at least at or about 93%, at least at or about 94%, at least at or about 95%, at or about 96%, at least at or about 97%, at least at or about 98%, or at least at or about 99% sequence identity to SEQ ID NO: 6 and retains fusogenic activity in conjunction with a variant NiV-F protein as described.

**[0152]** Reference to retaining fusogenic activity includes activity of a lipid particle (e.g. lentiviral vector) containing a variant NiV-F protein as described or biologically active portion or functionally active variant of the F protein (in conjunction with a G protein, such as a NiV-G protein as described) that is between at or about 10% and at or about 150% or more of the level or degree of binding of a reference lipid particle (e.g. lentiviral vector) that is similar, such as contains the same variant NiV-F, but that contains the corresponding wild-type G protein, such as set forth in SEQ ID NO: 5. For instance, a lipid particle (e.g. lentiviral vector) that retains fusogenic activity has at least or at least about 10% of the level or degree of fusogenic activity of the reference lipid particle that is similar (such as contains the same variant NiV-F) but that contains the corresponding wild-type G protein, such as at least or at least about 15% of the level or degree of fusogenic activity, at least or at least about 20% of the level or degree of fusogenic activity, at least or at least about 25% of the level or degree of fusogenic activity, at least or at least about 30% of the level or degree of fusogenic activity, at least or at least about 35% of the level or degree of fusogenic activity, at least or at least about 40% of the level or degree of fusogenic activity, at least or at least about 45% of the level or degree of fusogenic activity, at least or at least about 50% of the level or degree of fusogenic activity, at least or at least about 55% of the level or degree of fusogenic activity, at least or at least about 60% of the level or degree of fusogenic activity, at least or at least about 65% of the level or degree of fusogenic activity, at least or at least about 70% of the level or degree of fusogenic activity, at least or at least about 75% of the level or degree of fusogenic activity, at least or at least about 80% of the level or degree of fusogenic activity, at least or at least about 85% of the level or

degree of fusogenic activity, at least or at least about 90% of the level or degree of fusogenic activity, at least or at least about 95% of the level or degree of fusogenic activity, at least or at least about 100% of the level or degree of fusogenic activity, or at least or at least about 120% of the level or degree of fusogenic activity.

**[0153]** In some embodiments, the paramyxovirus envelope attachment protein is attached to a component of the universal adapter system (e.g., a tag polypeptide component or a catcher polypeptide component). In some embodiments, the paramyxovirus envelope attachment protein is attached to a component of the universal adapter system (e.g., a tag polypeptide component or a catcher polypeptide component) and the targeting moiety is fused to the other of the components (e.g., a tag polypeptide component or a catcher polypeptide component).

**[0154]** In some embodiments, the component of the universal adapter system is attached to the paramyxovirus envelope at any position, including in the globular head or stem (e.g., a NiV-G globular head or stem). In some embodiments, a component of the universal adapter system is attached to the C-terminus of the paramyxovirus envelope. The attachment to the C-terminus can be by direct fusion or indirect fusion, such as via a peptide linker. In some embodiments, the peptide linker is any as described herein, such as a GS linker of 4 to 12 amino acids.

**[0155]** In some embodiments, the component of the universal adapter system is attached via insertion or replacement of amino acid residues present in the paramyxovirus envelope attachment protein. In some embodiments, the component of the universal adapter system replaces any number of residues in a region of the paramyxovirus envelope attachment protein (e.g. NiV-G). The region can be a region of consecutive amino acids that form a loop between secondary structure in the protein. In some embodiments, the region is generally located on the protein's surface in an exposed area. It is within the level of a skilled artisan to determine appropriate regions, such as loops between secondary structures, in a paramyxovirus envelope protein in accord with the provided description. In some embodiments, the sequence of the component is inserted between amino acids in a region. In some embodiments, the sequence of the component is inserted by replacement of amino acids in a region. In some embodiments, the sequence of the component replaces all or a portion of the amino acid residues in a region.

**[0156]** With reference to the exemplary truncated NiV-G set forth in SEQ ID NO: 6, exemplary regions into which a component can be inserted are set forth in **Table 3A** and illustrated in **FIG. 7**. It is within the level of a skilled artisan to identify corresponding regions in other NiV-G envelope proteins such as by a sequence alignment between two similar protein sequences to identify corresponding residues.

**Table 3A**

| Site/ Region | SEQ ID NO | Sequence of Region | SEQ ID NO:6 | SEQ ID NO:5 |
|---|---|---|---|---|
| 1 | 21 | NLVGLPNNICLQKTSNQILKPKLI | 147-170 | 180-203 |
| 2 | 22 | PVVGQSGT | 175-182 | 208-215 |

| | | | | |
|---|---|---|---|---|
| 3 | 23 | MDEG | 191-194 | 224-227 |
| 4 | 24 | IGSCSRGV | 204-211 | 237-244 |
| 5 | 25 | DRGDEVP | 224-230 | 257-263 |
| 6 | 26 | WTPPNPNTV | 238-246 | 271-279 |
| 7 | 27 | NNEF | 254-257 | 287-290 |
| 8 | 28 | VSTVGDPILNSTYWSGSL | 264-281 | 297-314 |
| 9 | 29 | AVKPKSNG | 287-294 | 320-327 |
| 10 | 30 | LALRSIEKGRYDK | 302-314 | 335-347 |
| 11 | 31 | QGDT | 325-328 | 358-361 |
| 12 | 32 | VRTEFKYNDSNCPITKCQYSKP | 338-359 | 371-392 |
| 13 | 33 | MGIRPNSHY | 366-374 | 399-407 |
| 14 | 34 | NLSDGENPKV | 384-393 | 417-426 |
| 15 | 35 | EISDQRLSI | 397-405 | 430-438 |
| 16 | 36 | SLGQ | 414-417 | 447-450 |
| 17 | 37 | TVNPLV | 438-443 | 471-476 |
| 18 | 38 | GQSQCPRFNTCPAICAEG | 456-473 | 489-506 |
| 19 | 39 | DRINWI | 482-487 | 515-520 |
| 20 | 40 | LDSNATAAN | 493-501 | 526-534 |
| 21 | 41 | KDNEI | 508-512 | 541-545 |
| 22 | 42 | QLASEDTNA | 517-525 | 550-558 |
| 23 | 43 | LKNK | 535-538 | 568-571 |
| 24 | 44 | DTGDNV | 549-554 | 582-587 |
| 25 | 45 | KIPEQCT | 563-569 | 596-602 |

**[0157]** In particular embodiments, the G protein has the sequence of nucleic acids set forth in SEQ ID NO: 513, or is a functionally active variant thereof or a biologically active portion thereof that retains binding and/or fusogenic activity. In some embodiments, the G protein comprises a nucleic acid sequence having at least at or about 80%, at least at or about 85%, at least at or about 90%, at least at or about 91%, at least at or about 92%, at least at or about 93%, at least at or about 94%, at least at or about 95%, at or about 96%, at least at or about 97%, at least at or about 98%, or at least at or about 99% sequence identity to SEQ ID NO: 513.

**[0158]** In particular embodiments, the G protein has the sequence of nucleic acids set forth in SEQ ID NO: 514, or is a functionally active variant thereof or a biologically active portion thereof that retains binding and/or fusogenic activity. In some embodiments, the G protein comprises a nucleic acid sequence having at least at or about 80%, at least at or about 85%, at least at or about 90%, at least at or about 91%, at least at or about 92%, at least at or about 93%, at least at or about 94%, at least at or about 95%, at or about 96%, at least at or about 97%, at least at or about 98%, or at least at or about 99% sequence identity to SEQ ID NO: 514.

**[0159]** In particular embodiments, the G protein has the sequence of amino acids set forth in SEQ ID NO: 515, or is a functionally active variant thereof or a biologically active portion thereof that retains binding and/or fusogenic activity. In some embodiments, the G protein comprises an amino acid sequence having at least at or about 80%, at least at or about 85%, at least at or about 90%, at least at or about 91%, at least at or about 92%, at least at or about 93%, at least at or about 94%, at least at or about

95%, at or about 96%, at least at or about 97%, at least at or about 98%, or at least at or about 99% sequence identity to SEQ ID NO: 515.

**[0160]** In particular embodiments, the G protein has the sequence of amino acids set forth in SEQ ID NO: 516, or is a functionally active variant thereof or a biologically active portion thereof that retains binding and/or fusogenic activity. In some embodiments, the G protein comprises an amino acid sequence having at least at or about 80%, at least at or about 85%, at least at or about 90%, at least at or about 91%, at least at or about 92%, at least at or about 93%, at least at or about 94%, at least at or about 95%, at or about 96%, at least at or about 97%, at least at or about 98%, or at least at or about 99% sequence identity to SEQ ID NO: 516.

b. Mutated G proteins

**[0161]** In some embodiment, the G protein is a Paramyxovirus G glycoprotein (e.g., variant Paramyxovirus G glycoproteins) comprising one or more amino acid mutations that result in decreased glycosylation of the protein. The one or more amino acid mutations, also called deglycosylation mutations, can be one or more amino acid substitutions (also referred to as mutations).

**[0162]** In some embodiments, the mutant Paramyxovirus G glycoprotein comprises an amino acid substitution at one or more amino acid positions that reduce glycosylation of the G glycoprotein. In some embodiments, the one or more amino acid substitutions disrupts an N-linked glycosylation site. In some embodiments, the one or more amino acid substitutions disrupts an O-linked glycosylation site.

**[0163]** In some embodiments, the mutant Paramyxovirus G glycoprotein is derived from Morbillivirus (e.g., measles virus (MeV), canine distemper virus, Cetacean morbillivirus, Peste-des-petits-ruminants virus, Phocine distemper virus, Rinderpest virus), Henipavirus (e.g., Hendra (HeV) virus, Nipah (NiV) virus, a Cedar (CedPV) virus, Mòjiāng virus, a Langya virus or bat Paramyxovirus). In some embodiments, the mutant Paramyxovirus G glycoprotein is a mutant of a Paramyxovirus G glycoprotein derived from Nipah virus or Measles virus. In some embodiments, the mutant Paramyxovirus G protein is a mutant of a paramyxovirus G protein selected from the group consisting of SEQ ID NOs: 4, 5, 46-49, 252-257, 262-266, and 270-273 or a modified Paramyxovirus G glycoprotein derived from any one of 4, 5, 46-49, 252-257, 262-266, and 270-273 containing an altered cytoplasmic tail . In some embodiments, the mutant Paramyxovirus G protein has a sequence of amino acids that has at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94% or at least 95% to any one of SEQ ID NOs: 4, 5, 46-49, 252-257, 262-266, and 270-273 and contains the acid substitution at one or more amino acid positions that reduce glycosylation of the G glycoprotein as provided herein. In some embodiments, the mutant Paramyxovirus G protein that has one or more amino acid mutations that result in decreased glycosylation is a mutant of the truncated NiV-G set forth in SEQ ID NO:6.

**[0164]** The location of precited glycosylation sites can be determined using the sequence of a protein. For example, N-glycosylation often occurs at sites with the sequence N-X-S/T in which "X" is any amino acid except P. Various algorithms and tools are available for prediction of both N- and O-linked glycosylation, including SprintGly (http://sparks-lab.org/server/sprint-gly/), NetNGlyc (https://services.healthtech.dtu.dk/service.php?NetNGlyc-1.0), NetOGlyc (https://services.healthtech.dtu.dk/service.php?NetOGlyc-4.0), and GlycoMine$^{struct}$ (http://glycomine.erc.monash.edu/Lab/GlycoMine_Struct/), and methods described in Pitti et al., Sci. Reports, 9:15975 (2019) and Pakhrin et al., Molecules 26:7314 (2021). Any predicted glycosylation site may be substituted as described herein.

**[0165]** In some embodiments, the Paramyxovirus G glycoprotein to which the deglycosylation mutation is made is a NiV-G set forth in SEQ ID NO: 5 or a modified Nipah G glycoprotein (NiV-G) that has an altered cytoplasmic tail compared to native NiV-G (e.g., SEQ ID NO: 5). In some embodiments, the variant Paramyxovirus G protein has a sequence of amino acids that has at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94% or at least 95% to SEQ ID NO: 5 and contains the acid substitution at one or more amino acid positions that reduce glycosylation of the G glycoprotein as provided herein. In some embodiments, the Paramyxovirus G glycoprotein to which the deglycosylation mutation is made is a NiV-G set forth in SEQ ID NO: 5 or a modified Nipah G glycoprotein (NiV-G) that has an altered cytoplasmic tail compared to native NiV-G (e.g., SEQ ID NO: 5). In some embodiments, the variant Paramyxovirus G protein has a sequence of amino acids that has at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94% or at least 95% to SEQ ID NO: 5 and contains the acid substitution at one or more amino acid positions that reduce glycosylation of the G glycoprotein as provided herein.

**[0166]** Exemplary modified NiV-G proteins, including those with altered cytoplasmic tails, to which the one or more amino acid substitutions for reducing glycosylation can be incorporated are described in Section I.B.1.a.

**[0167]** Amino acid positions for substitutions are described herein with positions "corresponding to" positions of a reference sequence. It is understood that the amino acid substitutions are not limited to being made in only the reference sequence but also can be made in similar sequences by identification of residues that align or correspond with the reference positions. For instance, positions "corresponding to" to positions of a protein in a reference sequence can be identified upon alignment of a similar sequence with the referenced sequence based on structural sequence alignment or using a standard alignment algorithm, such as the GAP algorithm. By aligning the sequences, one skilled in the art can identify corresponding residues, for example, using conserved and identical amino acid residues as guides. For instance, amino acid positions for mutations are described herein with reference to the exemplary truncated NiV-G sequence set forth in SEQ ID NO:49; however, similar amino acid positions for

mutations as described can be made in other modified NiV-G sequences, such as any as described in Section I.B, by sequence alignment and identification of the corresponding residues.

**[0168]** In some embodiments, the one or more amino acid mutations are at positions corresponding to positions 39, 126, 128, 273, 345, 384, 448, and 496 of SEQ ID NO: 49. In some embodiments, the variant Paramyxovirus G glycoprotein comprises an amino acid mutation at any one of positions 39, 126, 128, 273, 345, 384, 448, and 496 of SEQ ID NO: 49. In some embodiments, the variant Paramyxovirus G glycoprotein comprises two or more amino acid mutations at any of positions corresponding to positions 39, 126, 128, 273, 345, 384, 448, and 496 of SEQ ID NO: 49, such as mutations at 2, 3, 4, 5, 7, or 8 of the positions.

**[0169]** In some embodiments, the one or more amino acid mutations is at a position corresponding to position 39 of SEQ ID NO: 49. In some embodiments, the one or more amino acid mutations is at a position corresponding to position 126 of SEQ ID NO: 49. In some embodiments, the one or more amino acid mutations is at a position corresponding to position 128 of SEQ ID NO: 49. In some embodiments, the one or more amino acid mutations is at a position corresponding to position 273 of SEQ ID NO: 49. In some embodiments, the one or more amino acid mutations is at a position corresponding to position 345 of SEQ ID NO: 49. In some embodiments, the one or more amino acid mutations is at a position corresponding to position 384 of SEQ ID NO: 49. In some embodiments, the one or more amino acid mutations is at a position corresponding to position 448 of SEQ ID NO: 49. In some embodiments, the one or more amino acid mutations is at a position corresponding to position 496 of SEQ ID NO: 49.

**[0170]** In some embodiments, the native amino acid at the position comprising the amino acid mutation is asparagine or serine. In some embodiments, the amino acid mutation is an amino acid substitution. In some embodiments, the mutation is an asparagine to glutamine substitution. In some embodiments, the mutation is a serine to alanine substitution.

**[0171]** In some embodiments, the mutation is an asparagine to glutamine substitution at a position corresponding to position 39 (N39Q) of SEQ ID NO: 49. In some embodiments, the mutation is an asparagine to glutamine substitution at a position corresponding to position 126 (N126Q) of SEQ ID NO: 49. In some embodiments, the mutation is an asparagine to glutamine substitution at a position corresponding to position 273 (N273Q) of SEQ ID NO: 49. In some embodiments, the mutation is an asparagine to glutamine substitution at a position corresponding to position 345 (N345Q) of SEQ ID NO: 49. In some embodiments, the mutation is an asparagine to glutamine substitution at a position corresponding to position 384 (N384Q) of SEQ ID NO: 49. In some embodiments, the mutation is an asparagine to glutamine substitution at a position corresponding to position 448 (N448Q) of SEQ ID NO: 49. In some embodiments, the mutation is an asparagine to glutamine substitution at a position corresponding to position 496 (N496Q) of SEQ ID NO: 49.

**[0172]** In some embodiments, the mutation is a serine to alanine substitution at a position corresponding to position 128 (S128A) of SEQ ID NO: 49.

**[0173]** In some embodiments, the G glycoprotein is derived from Nipah virus G protein and the one or more amino acid substitutions are at positions corresponding to positions selected from the group consisting of 39, 126, 128, 273, 345, 384, 448, and 496 of SEQ ID NO: 49. In some embodiments, the one or more amino acid substitutions are selected from N39Q, N126Q, S128A, N273Q, N345Q, N384Q, N448Q, N496Q or any combination thereof. In some embodiments, the G glycoprotein is a mutant NiV-G containing one amino acid substitution from any one of N39Q, N126Q, S128A, N273Q, N345Q, N384Q, N448Q, N496Q. In some embodiments, the G glycoprotein is a mutant NiV-G containing two amino acid substitutions from any two of N39Q, N126Q, S128A, N273Q, N345Q, N384Q, N448Q, N496Q. In some embodiments, the G glycoprotein is a mutant NiV-G containing three amino acid substitutions from any three of N39Q, N126Q, S128A, N273Q, N345Q, N384Q, N448Q, N496Q. In some embodiments, the G glycoprotein is a mutant NiV-G containing four amino acid substitutions from any one of N39Q, N126Q, S128A, N273Q, N345Q, N384Q, N448Q, N496Q. In some embodiments, the G glycoprotein is a mutant NiV-G containing five amino acid substitutions from any one of N39Q, N126Q, S128A, N273Q, N345Q, N384Q, N448Q, N496Q. In some embodiments, the G glycoprotein is a mutant NiV-G containing six amino acid substitutions from any one of N39Q, N126Q, S128A, N273Q, N345Q, N384Q, N448Q, N496Q. In some embodiments, the G glycoprotein is a mutant NiV-G containing seven amino acid substitutions from any one of N39Q, N126Q, S128A, N273Q, N345Q, N384Q, N448Q, N496Q. In some embodiments, the G glycoprotein is a mutant NiV-G containing eight amino acid substitutions from any one of N39Q, N126Q, S128A, N273Q, N345Q, N384Q, N448Q, N496Q. In some embodiments, the one or more amino acid substitutions are in the SEQ ID NO: 49 or a or a modified Nipah G glycoprotein (NiV-G) that has an altered cytoplasmic tail compared to native NiV-G (e.g., SEQ ID NO: 49). In some embodiments, the amino acid substitutions are in the NiV-G set forth in SEQ ID NO: 49.

**[0174]** In some embodiments, the variant Nipah-G protein comprises at least three amino acid substitutions. In some embodiments, the amino acid substitutions are at positions 273, 384, and 496 of SEQ ID NO:49. In some embodiments, the amino acid substitutions are at positions 273, 345, and 496 of SEQ ID NO:49. In some embodiments, the amino acid substitutions are at positions 39, 126, and 128 of SEQ ID NO:49. In some embodiments, the amino acid substitutions are at positions 39, 273, and 345 of SEQ ID NO:49. In some embodiments, the amino acid substitutions are at positions 39, 384, and 448 of SEQ ID NO:49. In some embodiments, the amino acid substitutions are at positions 39, 448, and 496 of SEQ ID NO:49. In some embodiments, the amino acid substitutions are at positions 39, 128, and 273 of SEQ ID NO:49. In some embodiments, the amino acid substitutions are at positions 39, 345, and 384 of SEQ ID NO:49. In some embodiments, the amino acid substitutions are at positions 39, 384, and 448 of SEQ ID NO:49.

**[0175]** In some embodiments, the variant Nipah-G protein comprises at least two amino acid substitutions. In some embodiments, the amino acid substitutions are at positions 273, and 496 of SEQ

ID NO:49. In some embodiments, the amino acid substitutions are at positions 345, and 496 of SEQ ID NO:49. In some embodiments, the amino acid substitutions are at positions 39 and 128 of SEQ ID NO:49. In some embodiments, the amino acid substitutions are at positions 39, and 345 of SEQ ID NO:49. In some embodiments, the amino acid substitutions are at positions 39, and 448 of SEQ ID NO:49. In some embodiments, the amino acid substitutions are at positions 39 and 496 of SEQ ID NO:49. In some embodiments, the amino acid substitutions are at positions 39 and 273 of SEQ ID NO:49. In some embodiments, the amino acid substitutions are at positions 39 and 384 of SEQ ID NO:49. In some embodiments, the amino acid substitutions are at positions 384 and 448 of SEQ ID NO:49.

**[0176]** In some embodiments, the amino acid substitution is at position 39 of SEQ ID NO:49. In some embodiments, the amino acid substitution is at position 126 of SEQ ID NO:49. In some embodiments, the amino acid substitution is at position 128 of SEQ ID NO:49. In some embodiments, the amino acid substitution is at position 273 of SEQ ID NO:49. In some embodiments, the amino acid substitution is at position 345 of SEQ ID NO:49. In some embodiments, the amino acid substitution is at position 384 of SEQ ID NO:49. In some embodiments, the amino acid substitution is at position 448 of SEQ ID NO:49. In some embodiments, the amino acid substitution is at position 496 of SEQ ID NO:49.

**[0177]** In some embodiments, the mutant Nipah-G protein comprises an asparagine to glutamine substitution at position 39 of SEQ ID NO:49. In some embodiments, the mutant Nipah-G protein comprises an asparagine to glutamine substitution at position 126 of SEQ ID NO:49. In some embodiments, the mutant Nipah-G protein comprises an asparagine to glutamine substitution at position 273 of SEQ ID NO:49. In some embodiments, the mutant Nipah-G protein comprises an asparagine to glutamine substitution at position 345 of SEQ ID NO:49. In some embodiments, the mutant Nipah-G protein comprises an asparagine to glutamine substitution at position 384 of SEQ ID NO:49. In some embodiments, the mutant Nipah-G protein comprises an asparagine to glutamine substitution at position 448 of SEQ ID NO:49. In some embodiments, the mutant Nipah-G protein comprises an asparagine to glutamine substitution at position 496 of SEQ ID NO:49. In some embodiments, the mutant Nipah-G protein comprises a serine to alanine substitution at position 128 of SEQ ID NO:49.

**[0178]** In some embodiments, the mutant Nipah-G protein comprises the sequence selected from the group consisting of any one of SEQ ID NOs: 381-507, such as any exemplary mutant Nipah-G proteins set forth in **Table 3B** below. In some embodiments, the mutant Nipah-G protein comprises the sequence of SEQ ID NO: 401. In some embodiments, the variant Nipah-G protein comprises the sequence of SEQ ID NO: 404. In some embodiments, the variant Nipah-G protein comprises the sequence of SEQ ID NO: 408.

**Table 3B: Exemplary Mutated Paramyxovirus Nipah G Proteins**

| NiV-G (Mutated) | SEQ ID NO | Position 39 | Position 126 | Position 273 | Position 345 | Position 384 | Position 448 | Position 496 |
|---|---|---|---|---|---|---|---|---|

| NivG | 49 | N | N | N | N | N | N | N |
|---|---|---|---|---|---|---|---|---|
| NivG.690 | 382 | Q | N | N | N | N | N | N |
| NivG.691 | 383 | N | Q | N | N | N | N | N |
| NivG.693 | 384 | N | N | Q | N | N | N | N |
| NivG.694 | 385 | N | N | N | Q | N | N | N |
| NivG.695 | 386 | N | N | N | N | Q | N | N |
| NivG.696 | 387 | N | N | N | N | N | Q | N |
| NivG.697 | 388 | N | N | N | N | N | N | Q |
| NivG.699 | 389 | N | N | N | Q | Q | Q | N |
| NivG.700 | 390 | N | N | N | Q | N | Q | N |
| NivG.701 | 391 | N | N | N | Q | N | N | Q |
| NivG.702 | 392 | Q | Q | Q | Q | Q | Q | Q |
| NivG.740 | 393 | N | N | N | N | N | Q | Q |
| NivG.742 | 394 | N | N | N | N | Q | N | Q |
| NivG.743 | 395 | N | N | N | N | Q | Q | N |
| NivG.744 | 396 | N | N | N | N | Q | Q | Q |
| NivG.748 | 397 | N | N | N | Q | N | Q | Q |
| NivG.749 | 398 | N | N | N | Q | Q | N | N |
| NivG.750 | 399 | N | N | N | Q | Q | N | Q |
| NivG.752 | 400 | N | N | N | Q | Q | Q | Q |
| NivG.753 | 401 | N | N | Q | N | N | N | Q |
| NivG.754 | 402 | N | N | Q | N | N | Q | N |
| NivG.755 | 403 | N | N | Q | N | N | Q | Q |
| NivG.756 | 404 | N | N | Q | N | Q | N | N |
| NivG.757 | 405 | N | N | Q | N | Q | N | Q |
| NivG.758 | 406 | N | N | Q | N | Q | Q | N |
| NivG.759 | 407 | N | N | Q | N | Q | Q | Q |
| NivG.760 | 408 | N | N | Q | Q | N | N | N |
| NivG.761 | 409 | N | N | Q | Q | N | N | Q |
| NivG.762 | 410 | N | N | Q | Q | N | Q | N |
| NivG.763 | 411 | N | N | Q | Q | N | Q | Q |
| NivG.764 | 412 | N | N | Q | Q | Q | N | N |
| NivG.765 | 413 | N | N | Q | Q | Q | N | Q |
| NivG.766 | 414 | N | N | Q | Q | Q | Q | N |
| NivG.767 | 415 | N | N | Q | Q | Q | Q | Q |
| NivG.768 | 416 | N | Q | N | N | N | N | Q |
| NivG.769 | 417 | N | Q | N | N | N | Q | N |
| NivG.770 | 418 | N | Q | N | N | N | Q | Q |
| NivG.771 | 419 | N | Q | N | N | Q | N | N |
| NivG.772 | 420 | N | Q | N | N | Q | N | Q |
| NivG.773 | 421 | N | Q | N | N | Q | Q | N |
| NivG.774 | 422 | N | Q | N | N | Q | Q | Q |
| NivG.775 | 423 | N | Q | N | Q | N | N | N |
| NivG.776 | 424 | N | Q | N | Q | N | N | Q |
| NivG.777 | 425 | N | Q | N | Q | N | Q | N |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| NivG.778 | 426 | N | Q | N | Q | N | Q | Q |
| NivG.779 | 427 | N | Q | N | Q | Q | N | N |
| NivG.780 | 428 | N | Q | N | Q | Q | N | Q |
| NivG.781 | 429 | N | Q | N | Q | Q | Q | N |
| NivG.782 | 430 | N | Q | N | Q | Q | Q | Q |
| NivG.783 | 431 | N | Q | Q | N | N | N | N |
| NivG.784 | 432 | N | Q | Q | N | N | N | Q |
| NivG.785 | 433 | N | Q | Q | N | N | Q | N |
| NivG.786 | 434 | N | Q | Q | N | N | Q | Q |
| NivG.787 | 435 | N | Q | Q | N | Q | N | N |
| NivG.788 | 436 | N | Q | Q | N | Q | N | Q |
| NivG.789 | 437 | N | Q | Q | N | Q | Q | N |
| NivG.790 | 438 | N | Q | Q | N | Q | Q | Q |
| NivG.791 | 439 | N | Q | Q | Q | N | N | N |
| NivG.792 | 440 | N | Q | Q | Q | N | N | Q |
| NivG.793 | 441 | N | Q | Q | Q | N | Q | N |
| NivG.794 | 442 | N | Q | Q | Q | N | Q | Q |
| NivG.795 | 443 | N | Q | Q | Q | Q | N | N |
| NivG.796 | 444 | N | Q | Q | Q | Q | N | Q |
| NivG.797 | 445 | N | Q | Q | Q | Q | Q | N |
| NivG.798 | 446 | N | Q | Q | Q | Q | Q | Q |
| NivG.799 | 447 | Q | N | N | N | N | N | Q |
| NivG.800 | 448 | Q | N | N | N | N | Q | N |
| NivG.801 | 449 | Q | N | N | N | N | Q | Q |
| NivG.802 | 450 | Q | N | N | N | Q | N | N |
| NivG.803 | 451 | Q | N | N | N | Q | N | Q |
| NivG.804 | 452 | Q | N | N | N | Q | Q | N |
| NivG.805 | 453 | Q | N | N | N | Q | Q | Q |
| NivG.806 | 454 | Q | N | N | Q | N | N | N |
| NivG.807 | 455 | Q | N | N | Q | N | N | Q |
| NivG.808 | 456 | Q | N | N | Q | N | Q | N |
| NivG.809 | 457 | Q | N | N | Q | N | Q | Q |
| NivG.810 | 458 | Q | N | N | Q | Q | N | N |
| NivG.811 | 459 | Q | N | N | Q | Q | N | Q |
| NivG.812 | 460 | Q | N | N | Q | Q | Q | N |
| NivG.813 | 461 | Q | N | N | Q | Q | Q | Q |
| NivG.814 | 462 | Q | N | Q | N | N | N | N |
| NivG.815 | 463 | Q | N | Q | N | N | N | Q |
| NivG.816 | 464 | Q | N | Q | N | N | Q | N |
| NivG.817 | 465 | Q | N | Q | N | N | Q | Q |
| NivG.818 | 466 | Q | N | Q | N | Q | N | N |
| NivG.819 | 467 | Q | N | Q | N | Q | N | Q |
| NivG.820 | 468 | Q | N | Q | N | Q | Q | N |
| NivG.821 | 469 | Q | N | Q | N | Q | Q | Q |
| NivG.822 | 470 | Q | N | Q | Q | N | N | N |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| NivG.823 | 471 | Q | N | Q | Q | N | N | Q |
| NivG.824 | 472 | Q | N | Q | Q | N | Q | N |
| NivG.825 | 473 | Q | N | Q | Q | N | Q | Q |
| NivG.826 | 474 | Q | N | Q | Q | Q | N | N |
| NivG.827 | 475 | Q | N | Q | Q | Q | N | Q |
| NivG.828 | 476 | Q | N | Q | Q | Q | Q | N |
| NivG.829 | 477 | Q | N | Q | Q | Q | Q | Q |
| NivG.830 | 478 | Q | Q | N | N | N | N | N |
| NivG.831 | 479 | Q | Q | N | N | N | N | Q |
| NivG.832 | 480 | Q | Q | N | N | N | Q | N |
| NivG.833 | 481 | Q | Q | N | N | N | Q | Q |
| NivG.834 | 482 | Q | Q | N | N | Q | N | N |
| NivG.835 | 483 | Q | Q | N | N | Q | N | Q |
| NivG.836 | 484 | Q | Q | N | N | Q | Q | N |
| NivG.837 | 485 | Q | Q | N | N | Q | Q | Q |
| NivG.838 | 486 | Q | Q | N | Q | N | N | N |
| NivG.839 | 487 | Q | Q | N | Q | N | N | Q |
| NivG.840 | 488 | Q | Q | N | Q | N | Q | N |
| NivG.841 | 489 | Q | Q | N | Q | N | Q | Q |
| NivG.842 | 490 | Q | Q | N | Q | Q | N | N |
| NivG.843 | 491 | Q | Q | N | Q | Q | N | Q |
| NivG.844 | 492 | Q | Q | N | Q | Q | Q | N |
| NivG.845 | 493 | Q | Q | N | Q | Q | Q | Q |
| NivG.846 | 494 | Q | Q | Q | N | N | N | N |
| NivG.847 | 495 | Q | Q | Q | N | N | N | Q |
| NivG.848 | 496 | Q | Q | Q | N | N | Q | N |
| NivG.849 | 497 | Q | Q | Q | N | N | Q | Q |
| NivG.850 | 498 | Q | Q | Q | N | Q | N | N |
| NivG.851 | 499 | Q | Q | Q | N | Q | N | Q |
| NivG.852 | 500 | Q | Q | Q | N | Q | Q | N |
| NivG.853 | 501 | Q | Q | Q | N | Q | Q | Q |
| NivG.854 | 502 | Q | Q | Q | Q | N | N | N |
| NivG.855 | 503 | Q | Q | Q | Q | N | N | Q |
| NivG.856 | 504 | Q | Q | Q | Q | N | Q | N |
| NivG.857 | 505 | Q | Q | Q | Q | N | Q | Q |
| NivG.858 | 506 | Q | Q | Q | Q | Q | N | N |
| NivG.859 | 507 | Q | Q | Q | Q | Q | N | Q |
| NivG.860 | 508 | Q | Q | Q | Q | Q | Q | N |

**[0179]** In some embodiments, the Paramyxovirus G glycoprotein to which the deglycosylation mutations is made is a Measles virus H (Mev-H) protein or a modified MeV-H protein that has an altered cytoplasmic tail compared to native MeV-H (e.g., SEQ ID NO:509). In some embodiments, the mutant Paramyxovirus G protein has a sequence of amino acids that has at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94% or at least

95% to SEQ ID NO: 509 and contains the acid substitution at one or more amino acid positions that reduce glycosylation of the G glycoprotein as provided herein.

**[0180]** In some embodiments, the G glycoprotein is derived from Measles virus H (Mev-H) protein and the one or more amino acid substitutions are at positions corresponding to positions selected from the group consisting of 168, 187, 200, 215, 238 of SEQ ID NO: 509. In some embodiments, the mutant Mev-H protein comprises at least two amino acid substitutions, such as 2, 3, 4, or 5 substitutions at positions 168, 187, 200, 215, 238 of SEQ ID NO: 509.

**[0181]** In some embodiments, the Paramyxovirus G glycoprotein to which the deglycosylation mutations is made is a Canine distemper virus H (CDV-H) protein or a modified CDV-H protein that has an altered cytoplasmic tail compared to native CDV-H (e.g., SEQ ID NO: 510). In some embodiments, the mutant Paramyxovirus G protein has a sequence of amino acids that has at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94% or at least 95% to SEQ ID NO: 510 and contains the acid substitution at one or more amino acid positions that reduce glycosylation of the G glycoprotein as provided herein.

**[0182]** In some embodiments, the G glycoprotein is derived from Canine distemper virus H (CDV-H) protein and the one or more amino acid substitutions are at positions corresponding to positions selected from the group consisting of 19, 149, 422 of SEQ ID NO: 510. In some embodiments, the variant CDV-H protein comprises at least two amino acid substitutions, such as 2 or 3 substitutions at positions 19, 149, 422 of SEQ ID NO: 510.

### *2. F protein*

**[0183]** In some embodiments, any of the provided lipid particles (lentiviral vectors) may also contain an F protein, such as a NiV-F protein, such as a full-length NiV-F protein or a biologically active portion thereof or a variant thereof. For instance, also provided herein are viral particles or viral-like particles, such as lentiviral particles or lentiviral-like particles, that are pseudotyped with any of the provided NiV-G proteins and a NiV-F protein, such as a full-length NiV-F protein or a biologically active portion or a variant thereof.

**[0184]** In some embodiments, the NiV-F protein exhibits fusogenic activity. In some embodiments, the NiV-F facilitates the fusion of the lipid particle (e.g. lentiviral vector) to a membrane. F proteins of henipaviruses, including NiV-F, are encoded as F0 precursors containing a signal peptide (e.g. corresponding to amino acid residues 1-26 of SEQ ID NO. 1). Following cleavage of the signal peptide, the mature F0 (SEQ ID NO:1 lacking the signal peptide, i.e. set forth in SEQ ID NO:2) is transported to the cell surface, then endocytosed and cleaved by cathepsin L (e.g. between amino acids 109-110 of NiV-F corresponding to amino acids set forth in SEQ ID NO:2) into the mature fusogenic subunits F1 (e.g. corresponding to amino acids 110-546 of NiV-F set forth in SEQ ID NO:2) and F2 (e.g. corresponding to amino acid residues 27-109 of NiV-F set forth in SEQ ID NO:2). The F1 and F2

subunits are associated by a disulfide bond and recycled back to the cell surface. The F1 subunit contains the fusion peptide domain located at the N terminus of the F1 subunit (e.g. corresponding to amino acids 110-129 of the below e.g. NiV-F set forth in SEQ ID NO:2) where it is able to insert into a cell membrane to drive fusion. In particular cases, fusion activity is blocked by association of the F protein with G protein, until G engages with a target molecule resulting in its disassociation from F and exposure of the fusion peptide to mediate membrane fusion.

**[0185]** Among different species, the sequence and activity of the F protein is highly conserved. For examples, the F protein of NiV and HeV viruses share 89% amino acid sequence identity. Further, in some cases, the henipavirus F proteins exhibit compatibility with G proteins from other species to trigger fusion (Brandel-Tretheway et al. Journal of Virology. 2019. 93(13):e00577-19). In some aspects of the provided lipid particles, the F protein is heterologous to the G protein, i.e. the F and G protein or biologically active portions are from different henipavirus species. For example, the F protein is from Hendra virus and the G protein is from Nipah virus. In other aspects, the F protein can be a chimeric F protein containing regions of F proteins from different species of Henipavirus. In some embodiments, switching a region of amino acid residues of the F protein from one species of Henipavirus to another can result in fusion to the G protein of the species comprising the amino acid insertion. (Brandel-Tretheway et al. 2019). In some cases, the chimeric F protein contains an extracellular domain from one henipavirus species and a transmembrane and/or cytoplasmic domain from a different henipavirus species. For example, the F protein contains an extracellular domain of Hendra virus and a transmembrane/cytoplasmic domain of Nipah virus. F protein sequences disclosed herein are predominantly disclosed as expressed sequences including an N-terminal signal sequence. As such N-terminal signal sequences are commonly cleaved co- or post-translationally, the mature protein sequences for all F protein sequences disclosed herein are also contemplated as lacking the N-terminal signal sequence.

**[0186]** In some embodiments, the F protein or the biologically active portion thereof is a wild-type Nipah virus F (NiV-F) protein or a Hendra virus F protein or is a functionally active variant or biologically active portion thereof. For instance, in some embodiments, the F protein or the biologically active portion thereof is a wild-type NiV-F protein or a functionally active variant or a biologically active portion thereof.

**[0187]** In some embodiments, the F protein has the sequence of amino acids set forth in SEQ ID NO:1, SEQ ID NO: 2, SEQ ID NO: 3, or SEQ ID NO: 4, or is a functionally active variant thereof or a biologically active portion thereof that retains fusogenic activity. In some embodiments, the functionally active variant comprises an amino acid sequence having at least at or about 80%, at least at or about 85%, at least at or about 90%, at least at or about 91%, at least at or about 92%, at least at or about 93%, at least at or about 94%, at least at or about 95%, at or about 96%, at least at or about 97%, at least at or about 98%, or at least at or about 99% sequence identity to SEQ ID NO:1, SEQ ID NO: 2, SEQ ID NO:

3, or SEQ ID NO. 4, and retains fusogenic activity in conjunction with a G protein, such as a NiV-G as provided herein. In some embodiments, the biologically active portion has an amino acid sequence having at least at or about 80%, at least at or about 85%, at least at or about 90%, at least at or about 91%, at least at or about 92%, at least at or about 93%, at least at or about 94%, at least at or about 95%, at or about 96%, at least at or about 97%, at least at or about 98%, or at least at or about 99% sequence identity to SEQ ID NO:1, SEQ ID NO: 2, SEQ ID NO: 3, or SEQ ID NO: 4.

**[0188]** In particular embodiments, the F protein has the sequence of amino acids set forth in SEQ ID NO:1, SEQ ID NO: 2, SEQ ID NO: 3, or SEQ ID NO: 4, or is a functionally active variant thereof or a biologically active portion thereof that retains fusogenic activity. In some embodiments, the functionally active variant comprises an amino acid sequence having at least at or about 80%, at least at or about 85%, at least at or about 90%, at least at or about 91%, at least at or about 92%, at least at or about 93%, at least at or about 94%, at least at or about 95%, at or about 96%, at least at or about 97%, at least at or about 98%, or at least at or about 99% sequence identity to SEQ ID NO:1, SEQ ID NO: 2, SEQ ID NO: 3, or SEQ ID NO: 4, and retains fusogenic activity in conjunction with a G protein, such as a NiV-G as provided herein. In some embodiments, the biologically active portion has an amino acid sequence having at least at or about 80%, at least at or about 85%, at least at or about 90%, at least at or about 91%, at least at or about 92%, at least at or about 93%, at least at or about 94%, at least at or about 95%, at or about 96%, at least at or about 97%, at least at or about 98%, or at least at or about 99% sequence identity to SEQ ID NO:1, SEQ ID NO: 2, SEQ ID NO: 3, or SEQ ID NO: 4.

**[0189]** Fusogenic activity includes the activity of the F protein in conjunction with a G protein to promote or facilitate fusion of two membrane lumens, such as the lumen of the targeted lipid particle having embedded in its lipid bilayer a henipavirus F and G protein, and a cytoplasm of a target cell, e.g. a cell that contains a surface receptor or molecule that is recognized or bound by the targeting moiety as fused to a component of a universal adapter system. In some embodiments, the F protein and G protein are from the same species (e.g. NiV-G and NiV-F). In some embodiments, the F protein and G protein are from different species (e.g. NiV-G and HeV-F). In particular embodiments, the F protein of the functionally active variant or biologically active portion retains the cleavage site cleaved by cathepsin L (e.g. corresponding to the cleavage site between amino acids 109-110 of SEQ ID NO:2).

**[0190]** Reference to retaining fusogenic activity includes activity (in conjunction with a G protein, such as a variant G protein provided herein) that is between at or about 10% and at or about 150% or more of the level or degree of binding of the corresponding wild-type F protein, such as set forth in SEQ ID NO:1, SEQ ID NO: 2, SEQ ID NO: 3, or SEQ ID NO: 4 or a cathepsin L cleaved from thereof containing an F1 and F2 subunit. In some embodiments, the fusogenic activity is at least or at least about 10% of the level or degree of fusogenic activity of the corresponding wild-type F protein, such as at least or at least about 15% of the level or degree of fusogenic activity of the corresponding wild-type F protein, such as at least or at least about 20% of the level or degree of fusogenic activity of the

corresponding wild-type F protein, such as at least or at least about 25% of the level or degree of fusogenic activity of the corresponding wild-type F protein, such as at least or at least about 30% of the level or degree of fusogenic activity of the corresponding wild-type F protein, such as at least or at least about 35% of the level or degree of fusogenic activity of the corresponding wild-type F protein, such as at least or at least about 40% of the level or degree of fusogenic activity of the corresponding wild-type F protein, such as at least or at least about 45% of the level or degree of fusogenic activity of the corresponding wild-type F protein, such as at least or at least about 50% of the level or degree of fusogenic activity of the corresponding wild-type F protein, such as at least or at least about 55% of the level or degree of fusogenic activity of the corresponding wild-type F protein, such as at least or at least about 60% of the level or degree of fusogenic activity of the corresponding wild-type F protein, such as at least or at least about 65% of the level or degree of fusogenic activity of the corresponding wild-type F protein, such as at least or at least about 70% of the level or degree of fusogenic activity of the corresponding wild-type F protein, such as at least or at least about 75% of the level or degree of fusogenic activity of the corresponding wild-type F protein, such as at least or at least about 80% of the level or degree of fusogenic activity of the corresponding wild-type F protein, such as at least or at least about 85% of the level or degree of fusogenic activity of the corresponding wild-type F protein, such as at least or at least about 90% of the level or degree of fusogenic activity of the corresponding wild-type F protein, such as at least or at least about 95% of the level or degree of fusogenic activity of the corresponding wild-type F protein, such as at least or at least about 100% of the level or degree of fusogenic activity of the corresponding wild-type F protein, or such as at least or at least about 120% of the level or degree of fusogenic activity of the corresponding wild-type F protein.

**[0191]** In some embodiments, the mutant F protein is a biologically active portion that is truncated and lacks up to 22 contiguous amino acid residues at or near the C-terminus of the wild-type F protein, such as a wild-type F protein set forth in any one of SEQ ID NO:1 or 2. In some embodiments, the mutant F protein is truncated and lacks up to 22 contiguous amino acids, such as up to 21, 20, 19, 18 , 17, 16, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2, or 1 contiguous amino acids at the C-terminus of the wild-type F protein.

**[0192]** In some embodiments, the NiV-F, such as a mutant or truncated NiV-F, of a provided lipid particle includes the F0 precursor or a proteolytically cleaved form thereof containing the F1 and F2 subunits, such as resulting following proteolytic cleavage at the cleavage site (e.g. between amino acids corresponding to amino acids between amino acids 109-110 of SEQ ID NO:2) to produce two chains that can be linked by disulfide bond. In some embodiments, the NiV-F, such as wild-type NiV-F or a truncated or mutated NiV- F protein, is produced or encoded as an F0 precursor which then is able to be proteolytically cleaved to result in an F protein containing the F1 and F2 subunit linked by a disulfide bond. Hence, it is understood that reference to a particular sequence (SEQ ID NO) of a NiV-F herein is typically with reference to the F0 precursor sequence but also is understood to include the proteolytically

cleaved form or sequence thereof containing the two cleaved chains, F1 and F2. For instance, the NiV-F, such as a mutant or truncated NiV-F, contains an F1 subunit corresponding to amino acids 110-546 of NiV-F set forth in SEQ ID NO:2 or truncated or mutant sequence thereof, and an F2 corresponding to amino acid residues 27-109 of NiV-F set forth in SEQ ID NO:2.

**[0193]** In some embodiments, the mutant F protein is a biologically active portion that is truncated and lacks up to 22 contiguous amino acid residues at or near the C-terminus of the wild-type NiV-F protein, such as a wild-type NiV-F protein set forth in SEQ ID NO:1 or SEQ ID NO:2. In some embodiments, the mutant F protein is truncated and lacks up to 22 contiguous amino acids, such as up to 21, 20, 19, 18 , 17, 16, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2, or 1 contiguous amino acids at the C-terminus of the wild-type NiV-F protein, such as a wild-type NiV-F protein set forth in SEQ ID NO:1 or SEQ ID NO:2. In some embodiments, the mutant F protein contains an F1 subunit and an F2 subunit in which (1) the F1 subunit is truncated and lacks up to 22 contiguous amino acids at or near the C-terminus of the wild-type F1 subunit, such as lacks up to 22 contiguous amino acids at or near the C-terminus of the wild-type F1 subunit corresponding to amino acids 110-546 of NiV-F set forth in SEQ ID NO:2, and (2) the F2 subunit has the sequence corresponding to amino acid residues 27-109 of NiV-F set forth in SEQ ID NO:2.

**[0194]** In some embodiments, the F protein is a mutant NiV-F protein that is a biologically active portion thereof that comprises a 22 amino acid truncation at or near the C-terminus of the wild-type NiV-F protein (SEQ ID NO:1 or SEQ ID NO:2). In some embodiments, the NiV-F protein is encoded by a nucleotide sequence that encodes the sequence set forth in SEQ ID NO: 3. In some embodiments, the NiV-F proteins is encoded by a nucleotide sequence that encodes sequence having at least at or about 90%, at least at or about 91%, at least at or about 92%, at least at or about 93%, at least at or about 94%, at least at or about 95%, at or about 96%, at least at or about 97%, at least at or about 98%, or at least at or about 99% sequence identity to SEQ ID NO: 3. In particular embodiments, the variant F protein is a mutant Niv-F protein that has the sequence of amino acids set forth in SEQ ID NO:4. In some embodiments, the NiV-F proteins is encoded by a sequence having at least at or about 90%, at least at or about 91%, at least at or about 92%, at least at or about 93%, at least at or about 94%, at least at or about 95%, at or about 96%, at least at or about 97%, at least at or about 98%, or at least at or about 99% sequence identity to SEQ ID NO: 4. In some embodiments, the F protein molecule or biologically active portion thereof comprises the sequence set forth in SEQ ID NO: 4.

### C. Targeting Moiety

**[0195]** In some embodiments, a paramyxovirus envelope attachment protein such as any disclosed herein is attached to a targeting moiety via a universal adapter system. In some embodiments, a paramyxovirus envelope attachment protein is attached to a targeting moiety via a universal adapter system, wherein the paramyxovirus envelope attachment protein is fused to one of the components, and

the targeting moiety is fused to the other of the components. In some embodiments, the paramyxovirus envelope attachment protein is fused to the tag polypeptide component of a universal adapter system, and the targeting moiety is fused to the catcher polypeptide component of a universal adapter system. In some embodiments, the paramyxovirus envelope attachment protein is fused to the catcher polypeptide component of a universal adapter system, and the targeting moiety is fused to the tag polypeptide component of a universal adapter system.

**[0196]** In some embodiments, provided herein is a paramyxovirus envelope attachment protein with a targeting moiety attached via a universal adapter system, said system comprising a tag polypeptide component and a catcher polypeptide component, wherein the paramyxovirus envelope attachment protein is fused to the tag polypeptide component, and the targeting moiety is fused to the catcher polypeptide component. In some embodiments, the tag polypeptide component fused to the paramyxovirus envelope attachment protein is linked to the C-terminus of the paramyxovirus envelope attachment protein. In some embodiments, the tag polypeptide component fused to the paramyxovirus envelope attachment protein is inserted into a loop region of the paramyxovirus envelope attachment protein.

**[0197]** In some embodiments, provided herein is a paramyxovirus envelope attachment protein with a targeting moiety attached via a universal adapter system, said system comprising a tag polypeptide component and a catcher polypeptide component, wherein the paramyxovirus envelope attachment protein is fused to the catcher polypeptide component, and the targeting moiety is fused to the tag polypeptide component. In some embodiments, the catcher polypeptide component fused to the paramyxovirus envelope attachment protein is inserted into a loop region of the paramyxovirus envelope attachment protein. In some embodiments, the catcher polypeptide component fused to the paramyxovirus envelope attachment protein is linked to the C-terminus of the paramyxovirus envelope attachment protein.

**[0198]** In some embodiments, the targeting moiety can be any agent that binds to a cell surface molecule on a target cell. In some embodiments, the targeting moiety can be an antibody or an antibody portion or fragment.

**[0199]** The targeting moiety may be modulated to have different binding strengths. For example, scFvs and antibodies with various binding strengths may be used to alter the fusion activity of the chimeric attachment proteins towards cells that display high or low amounts of the target antigen. For example DARPins with different affinities may be used to alter the fusion activity towards cells that display high or low amounts of the target antigen. Targeting moieties may also be modulated to target different regions on the target ligand, which will affect the fusion rate with cells displaying the target.

**[0200]** The targeting moiety may comprise a humanized antibody molecule, intact IgA, IgG, IgE or IgM antibody; bi- or multi- specific antibody (e.g., Zybodies®, etc.); antibody fragments such as

Fab fragments, Fab' fragments, F(ab')2 fragments, Fd' fragments, Fd fragments, and isolated CDRs or sets thereof; single chain Fvs; polypeptide-Fc fusions; single domain antibodies (e.g., shark single domain antibodies such as IgNAR or fragments thereof); cameloid antibodies; masked antibodies (e.g., Probodies®); Small Modular ImmunoPharmaceuticals ("SMIPsTM"); single chain or Tandem diabodies (TandAb®); VHHs; Anticalins®; Nanobodies®; minibodies; BiTE®s; ankyrin repeat proteins or DARPINs®; Avimers®; DARTs; TCR-like antibodies;, Adnectins®; Affilins®; Trans-bodies®; Affibodies®; TrimerX®; MicroProteins; Fynomers®, Centyrins®; and KALBITOR®s. A targeting moiety can also include an antibody or an antigen-binding fragment thereof (e.g., Fab, Fab', F(ab')2, Fv fragments, scFv antibody fragments, disulfide-linked Fvs (sdFv), a Fd fragment consisting of the VH and CH1 domains, linear antibodies, single domain antibodies such as sdAb (either VL or VH), nanobodies, or camelid VHH domains), an antigen-binding fibronectin type III (Fn3) scaffold such as a fibronectin polypeptide minibody, a ligand, a cytokine, a chemokine, or a T cell receptor (TCRs).

**[0201]** In some embodiments, the targeting moiety is a single chain molecule. In some embodiments, the targeting moiety is a single domain antibody. In some embodiments, the targeting moiety is a single chain variable fragment. In particular embodiments, the targeting moiety contains an antibody variable sequence that is human or humanized.

**[0202]** In some embodiments, the targeting moiety is a single domain antibody. In some embodiments, the single domain antibody can be human or humanized. In some embodiments, the single domain antibody or portion thereof is naturally occurring. In some embodiments, the single domain antibody or portion thereof is synthetic.

**[0203]** In some embodiments, the single domain antibodies are antibodies whose complementary determining regions are part of a single domain polypeptide. In some embodiments, the single domain antibody is a heavy chain only antibody variable domain. In some embodiments, the single domain antibody does not include light chains.

**[0204]** In some embodiments, the heavy chain antibody devoid of light chains is referred to as VHH. In some embodiments, the single domain antibody antibodies have a molecular weight of 12-15 kDa. In some embodiments, the single domain antibody antibodies include camelid antibodies or shark antibodies. In some embodiments, the single domain antibody molecule is derived from antibodies raised in Camelidae species, for example in camel, llama, dromedary, alpaca, vicuna and guanaco. In some embodiments, the single domain antibody is referred to as immunoglobulin new antigen receptors (IgNARs) and is derived from cartilaginous fishes. In some embodiments, the single domain antibody is generated by splitting dimeric variable domains of human or mouse IgG into monomers and camelizing critical residues.

**[0205]** In some embodiments, the single domain antibody can be generated from phage display libraries. In some embodiments, the phage display libraries are generated from a VHH repertoire of

camelids immunized with various antigens, as described in Arbabi et al., *FEBS Letters,* 414, 521-526 (1997); Lauwereys et al., *EMBO J.,* 17, 3512-3520 (1998); Decanniere et al., *Structure,* 7, 361-370 (1999). In some embodiments, the phage display library is generated comprising antibody fragments of a non-immunized camelid. In some embodiments, single domain antibodies a library of human single domain antibodies is synthetically generated by introducing diversity into one or more scaffolds. In some embodiments, the single chain antibody is an scFv.

**[0206]** In some embodiments, a targeting moiety (e.g. sdAb or one of any targeting moieties as described herein) binds to a cell surface antigen of a cell. In some embodiments, a cell surface antigen is characteristic of one type of cell. In some embodiments, a cell surface antigen is characteristic of more than one type of cell.

**[0207]** In some embodiments, the cell surface molecule of a target cell is an antigen or portion thereof. In some embodiments, the single domain antibody or portion thereof is an antibody having a single monomeric domain antigen binding/recognition domain that is able to bind selectively to a specific antigen. In some embodiments, the single domain antibody binds an antigen present on a target cell.

**[0208]** Exemplary cells include polymorphonuclear cells (also known as PMN, PML, PMNL, or granulocytes), stem cells, embryonic stem cells, neural stem cells, mesenchymal stem cells (MSCs), hematopoietic stem cells (HSCs), human myogenic stem cells, muscle-derived stem cells (MuStem), embryonic stem cells (ES or ESCs), limbal epithelial stem cells, cardio-myogenic stem cells, cardiomyocytes, progenitor cells, immune effector cells, lymphocytes, macrophages, dendritic cells, natural killer cells, T cells, cytotoxic T lymphocytes, allogenic cells, resident cardiac cells, induced pluripotent stem cells (iPS), adipose-derived or phenotypic modified stem or progenitor cells, CD133+ cells, aldehyde dehydrogenase-positive cells (ALDH+), umbilical cord blood (UCB) cells, peripheral blood stem cells (PBSCs), neurons, neural progenitor cells, pancreatic beta cells, glial cells, or hepatocytes,

**[0209]** In some embodiments, the target cell is a cell of a target tissue. The target tissue can include liver, lungs, heart, spleen, pancreas, gastrointestinal tract, kidney, testes, ovaries, brain, reproductive organs, central nervous system, peripheral nervous system, skeletal muscle, endothelium, inner ear, or eye.

**[0210]** In some embodiments, the target cell is a muscle cell (e.g., skeletal muscle cell), kidney cell, liver cell (e.g. hepatocyte), or a cardiac cell (e.g. cardiomyocyte). In some embodiments, the target cell is a cardiac cell, e.g., a cardiomyocyte (e.g., a quiescent cardiomyocyte), a hepatoblast (e.g., a bile duct hepatoblast), an epithelial cell, a T cell (e.g. a naive T cell), a macrophage (e.g., a tumor infiltrating macrophage), or a fibroblast (e.g., a cardiac fibroblast).

**[0211]** In some embodiments, the target cell is a tumor-infiltrating lymphocyte, a T cell, a neoplastic or tumor cell, a virus-infected cell, a stem cell, a central nervous system (CNS) cell, a hematopoeietic stem cell (HSC), a liver cell or a fully differentiated cell. In some embodiments, the

target cell is a CD3+ T cell, a CD4+ T cell, a CD8+ T cell, a hepatocyte, a hematopoietic stem cell, a CD34+ hematopoietic stem cell, a CD105+ hematopoietic stem cell, a CD117+ hematopoietic stem cell, a CD105+ endothelial cell, a B cell, a CD20+ B cell, a CD19+ B cell, a cancer cell, a CD133+ cancer cell, an EpCAM+ cancer cell, a CD19+ cancer cell, a Her2/Neu+ cancer cell, a GluA2+ neuron, a GluA4+ neuron, a NKG2D+ natural killer cell, a SLC1A3+ astrocyte, a SLC7A10+ adipocyte, or a CD30+ lung epithelial cell.

**[0212]** In some embodiments, the target cell is an antigen presenting cell, an MHC class II+ cell, a professional antigen presenting cell, an atypical antigen presenting cell, a macrophage, a dendritic cell, a myeloid dendritic cell, a plasmacyteoid dendritic cell, a CD11c+ cell, a CD11b+ cell, a splenocyte, a B cell, a hepatocyte, a endothelial cell, or a non-cancerous cell).

**[0213]** In some embodiments, the targeting moiety (e.g. sdAb) variable domain binds a cell surface molecule or antigen. In some embodiments, the cell surface molecule is ASGR1, ASGR2, TM4SF5, CD8, CD4, or low density lipoprotein receptor (LDL-R). In some embodiments, the cell surface molecule is ASGR1. In some embodiments, the cell surface molecule is ASGR2. In some embodiments, the cell surface molecule is TM4SF5. In some embodiments, the cell surface molecule is CD8. In some embodiments, the cell surface molecule is CD4. In some embodiments, the cell surface molecule is LDL-R.

**[0214]** In some embodiments, the target molecule is a cell surface molecule selected from the group consisting of CD3, CD4, CD7 CD8, ASCT2, CD105, CD110, CD117, CD133, CD146, CD164, CD34, CD46, CD49f, CD90, EPCR, and ITGA3.

**[0215]** In some embodiments, the targeting moiety targets a target molecule on the surface of a hepatocyte. In some embodiments, the target molecule is a cell surface molecule selected from the group consisting of ASGR1, ASGR2 and TM4SF5.

**[0216]** In some embodiments, the targeting moiety targets a target molecule on the surface of a T cell. In some embodiments, the target molecule is CD3, CD4 or CD8.

## II. LIPID PARTICLES COMPRISING A UNIVERSAL ADAPATER SYSTEM AND METHODS OF PRODUCTION

**[0217]** Provided herein is a lipid particle comprising a lipid bilayer, a lumen surrounded by the lipid bilayer and a variant NiV-G protein, such as any as described, in which the variant NiV-G is embedded within the lipid bilayer. In some embodiments, the provided lipid particles preferentially target hematopoietic cells (e.g. T cells), which is mediated by the tropism of the variant NiV-G. In some embodiments, the lipid particle may additionally contain an exogenous agent (e.g. therapeutic agent) for delivery to a cell. In some embodiments, a lipid particle is introduced to a cell in the subject. Also provided are methods of delivering any of the provided lipid particles to a cell.

**[0218]** In some embodiments, the provided lipid particles exhibit fusogenic activity, which is mediated by the variant NiV-G along with any of the provided F proteins that facilitates merger or fusion of the two lumens of the lipid particle and the target cell membranes. Thus, among provided lipid particles are fusosomes. In some embodiments, the fusosome comprises a naturally derived bilayer of amphipathic lipids with the variant NiV-G as a fusogen. In some embodiments, the fusosome comprises (a) a lipid bilayer, (b) a lumen (e.g., comprising cytosol) surrounded by the lipid bilayer; and (c) a fusogen that is exogenous or overexpressed relative to the source cell. In some embodiments, the variant NiV-G disposed in the lipid bilayer. In some embodiments, the fusosome comprises several different types of lipids, e.g., amphipathic lipids, such as phospholipids

**[0219]** In some embodiments, the lipid particle includes a naturally derived bilayer of amphipathic lipids that encloses lumen or cavity. In some embodiments, the lipid particle comprises a lipid bilayer as the outermost surface. In some embodiments, the lipid bilayer encloses a lumen. In some embodiments, the lumen is aqueous. In some embodiments, the lumen is in contact with the hydrophilic head groups on the interior of the lipid bilayer. In some embodiments, the lumen is a cytosol. In some embodiments, the cytosol contains cellular components present in a source cell. In some embodiments, the cytosol does not contain components present in a source cell. In some embodiments, the lumen is a cavity. In some embodiments, the cavity contains an aqueous environment. In some embodiments, the cavity does not contain an aqueous environment.

**[0220]** In some aspects, the lipid bilayer is derived from a source cell during a process to produce a lipid-containing particle. Exemplary methods for producing lipid-containing particles are provided in Section II.A.3. In some embodiments, the lipid bilayer includes membrane components of the cell from which the lipid bilayer is produced, e.g., phospholipids, membrane proteins, etc. In some embodiments, the lipid bilayer includes a cytosol that includes components found in the cell from which the micro-vesicle is produced, e.g., solutes, proteins, nucleic acids, etc., but not all of the components of a cell, e.g., they lack a nucleus. In some embodiments, the lipid bilayer is considered to be exosome-like. The lipid bilayer may vary in size, and in some instances have a diameter ranging from 30 and 300 nm, such as from 30 and 150 nm, and including from 40 to 100 nm.

**[0221]** In some embodiments, the lipid bilayer is a viral envelope. In some embodiments, the viral envelope is obtained from a source cell. In some embodiments, the viral envelope is obtained by the viral capsid from the source cell plasma membrane. In some embodiments, the lipid bilayer is obtained from a membrane other than the plasma membrane of a host cell. In some embodiments, the viral envelope lipid bilayer is embedded with viral proteins, including viral glycoproteins as described herein such as a variant NiV-G protein and, in some aspects, also a NiV-F protein.

**[0222]** In other aspects, the lipid bilayer includes synthetic lipid complex. In some embodiments, the synthetic lipid complex is a liposome. In some embodiments, the lipid bilayer is a vesicular structure characterized by a phospholipid bilayer membrane and an inner aqueous medium. In some embodiments,

the lipid bilayer has multiple lipid layers separated by aqueous medium. In some embodiments, the lipid bilayer forms spontaneously when phospholipids are suspended in an excess of aqueous solution. In some examples, the lipid components undergo self-rearrangement before the formation of closed structures and entrap water and dissolved solutes between the lipid bilayers.

**[0223]** In some embodiments, a targeted envelope protein and fusogen, such as any described above including any that are exogenous or overexpressed relative to the source cell, is disposed in the lipid bilayer.

**[0224]** In some embodiments, the lipid particle comprises several different types of lipids. In some embodiments, the lipids are amphipathic lipids. In some embodiments, the amphipathic lipids are phospholipids. In some embodiments, the phospholipids comprise phosphatidylcholine, phosphatidylethanolamine, phosphatidylinositol, and phosphatidylserine. In some embodiments, the lipids comprise phospholipids such as phosphocholines and phosphoinositols. In some embodiments, the lipids comprise DMPC, DOPC, and DSPC.

**[0225]** In some embodiments, the bilayer may be comprised of one or more lipids of the same or different type. In some embodiments, the source cell comprises a cell selected from HEK293 cells, CHO cells, BHK cells, MDCK cells, C3H 10T1/2 cells, FLY cells, Psi-2 cells, BOSC 23 cells, PA317 cells, WEHI cells, COS cells, BSC 1 cells, BSC 40 cells, BMT 10 cells, VERO cells, W138 cells, MRC5 cells, A549 cells, HT1080 cells, 293 cells, 293T cells, B-50 cells, 3T3 cells, NIH3T3 cells, HepG2 cells, Saos-2 cells, Huh7 cells, HeLa cells, W163 cells, 211 cells, and 211A cells.

**[0226]** In some embodiments, the lipid particle can be a viral particle, a virus-like particle, a nanoparticle, a vesicle, an exosome, a dendrimer, a lentivirus, a viral vector, an enucleated cell, a microvesicle, a membrane vesicle, an extracellular membrane vesicle, a plasma membrane vesicle, a giant plasma membrane vesicle, an apoptotic body, a mitoparticle, a pyrenocyte, a lysosome, another membrane enclosed vesicle, or a lentiviral vector, a viral based particle, a virus like particle (VLP) or a cell based particle.

**[0227]** In particular embodiments, the lipid particle is virally derived. In some embodiments, the lipid particle can be a viral-based particle, such as a viral vector particle (e.g. lentiviral vector particle) or a virus-like particle (e.g. a lentiviral-like particle). In some embodiments, the lipid bilayer is a viral envelope. In some embodiments, the viral envelope is obtained from a host cell. In some embodiments, the viral envelope is obtained by the viral capsid from the source cell plasma membrane. In some embodiments, the lipid bilayer is obtained from a membrane other than the plasma membrane of a host cell. In some embodiments, the viral envelope lipid bilayer is embedded with viral proteins, including viral glycoproteins.

**[0228]** In particular embodiments, the lipid particle is not virally derived. In some embodiments, the lipid particle can be a nanoparticle, a vesicle, an exosome, a dendrimer, an enucleated cell, a microvesicle, a membrane vesicle, an extracellular membrane vesicle, a plasma membrane vesicle, a

giant plasma membrane vesicle, an apoptotic body, a mitoparticle, a pyrenocyte, a lysosome, another membrane enclosed vesicle, or a cell derived particle.

**[0229]** In some embodiments, the lipid bilayer includes membrane components of the host cell from which the lipid bilayer is derived, e.g., phospholipids, membrane proteins, etc. In some embodiments, the lipid bilayer includes a cytosol that includes components found in the cell from which the vehicle is derived, e.g., solutes, proteins, nucleic acids, etc., but not all of the components of a cell, e.g., lacking a nucleus. In some embodiments, the lipid bilayer is considered to be exosome-like. The lipid bilayer may vary in size, and in some instances have a diameter ranging from 30 and 300 nm, such as from 30 and 150 nm, and including from 40 to 100 nm.

**[0230]** In particular embodiments, an exogenous agent, such as a polynucleotide or polypeptide, is encapsulated within the lumen of a lipid particle. Embodiments of provided lipid particles may have various properties that facilitate delivery of a payload, such as, e.g., a desired transgene or exogenous agent, to a target cell. The exogenous agent may be a polynucleotide or a polypeptide. In some embodiments, a lipid particle provided herein is administered to a subject, e.g., a mammal, e.g., a human. In such embodiments, the subject may be at risk of, may have a symptom of, or may be diagnosed with or identified as having, a particular disease or condition. In one embodiment, the subject has cancer. In one embodiment, the subject has an infectious disease. In some embodiments, the lipid particle contains nucleic acid sequences (polynucleotide) encoding an exogenous agent or a polypeptide exogenous agent for treating the disease or condition.

**[0231]** The lipid particles can include spherical particles or can include particles of elongated or irregular shape.

**[0232]** In some embodiments, a composition of particles can be assessed for one or more features related to their size, including diameter, range of variation thereof above and below an average (mean) or median value of the diameter, coefficient of variation, polydispersity index or other measure of size of particles in a composition. Various methods for particle characterization can be used, including, but not limited to, laser diffraction, dynamic light scattering (DLS; also known as photon correlation spectroscopy) or image analysis, such as microscopy or automated image analysis.

**[0233]** In some embodiments, the provided lipid particle has a diameter of, or the average (mean) diameter of particles in a composition is, less than about 3 μm, less than about 2 μm, less than about 1 μm, less than about 900 nm, less than about 800 nm, less than about 700 nm, less than about 600 nm, less than about 500 m, less than about 400 nm, less than about 300, less than about 200 nm, less than about 150 nm, less than about 100 nm, less than about 50 nm, or less than about 20 nm. In some embodiments, the lipid particle has a diameter of, or the average (mean) diameter of particles in a composition is, less than about 400 nm. In another embodiment, the lipid particle has a diameter of, or the average (mean) diameter of particles in a composition is, less than about 150 nm. In some embodiments, the lipid particle has a diameter of, or the average (mean) diameter of particles in a composition is, between at or about 2

μm and at or about 1 μm, between at or about 1 μm and at or about 900 nm, between at or about 900 nm and at or about 800 nm, between at or about 800 and at or about 700 nm, between at or about 700 nm and at or about 600 nm, between at or about 600 nm and at or about 500 nm, between at or about 500 nm and at or about 400 nm, between at or about 400 nm and at or about 300 nm, between at or about 300 nm and at or about 200 nm, between at or about 200 and at or about 100 nm, between at or about 100 and at or about 50 nm, or between at or about 20 nm and at or about 50 nm.

**[0234]** In some embodiments the median particle diameter in a composition of particles is between at or about 10 nm and at or about 1000 nM, between at or about 25 nm and at or about 500 nm, between at or about 40 nm and at or about 300 nm, between at or about 50 nm and at or about 250 nm, between at or about 60 nm and at or about 225 nm, between at or about 70 nm and at or about 200 nm, between at or about 80 nm and at or about 175 nm, or between at or about 90 nm and at or about 150 nm.

**[0235]** In some embodiments, 90% of the lipid particles in a composition fall within 50% of the median diameter of the lipid particles. In some embodiments, 90% of the lipid particles in a composition fall within 25% of the median diameter of the lipid particles. In some embodiments, 90% of the lipid particles in a composition fall within 20% of the median diameter. In some embodiments, 90% of the lipid particles in a composition fall within 15% of the median diameter of lipid particles. In some embodiments, 90% of the lipid particles in a composition fall within 10% of the median diameter of the lipid particles.

**[0236]** In some embodiments, 75% of the lipid particles in a composition fall within +/- 2 or +/- 1 St Dev standard deviations (St Dev) of the mean diameter of lipid particles. In some embodiments, 80% of the lipid particles in a composition fall within +/- 2 St Dev or +/- 1 St Dev of the mean diameter of lipid particles. In some embodiments, 85% of the lipid particles in a composition fall within +/- 2 St Dev or +/- 1 St Dev of the mean diameter of lipid particles. In some embodiments, 90% of the lipid particles in a composition fall within +/- 2 St Dev or +/- 1 St Dev of the mean diameter of lipid particles. In some embodiments, 95% of the lipid particles in a composition fall within +/- 2 St Dev or +/- 1 St Dev of the mean diameter of lipid particles.

**[0237]** In some embodiments, the lipid particles have an average hydrodynamic radius, e.g. as determined by DLS, of about 100 nm to about two microns. In some embodiments, the lipid particles have an average hydrodynamic radius between at or about 2 μm and at or about 1 μm, between at or about 1 μm and at or about 900 nm, between at or about 900 nm and at or about 800 nm, between at or about 800 and at or about 700 nm, between at or about 700 nm and at or about 600 nm, between at or about 600 nm and at or about 500 nm, between at or about 500 nm and at or about 400 nm, between at or about 400 nm and at or about 300 nm, between at or about 300 nm and at or about 200 nm, between at or about 200 and at or about 100 nm, between at or about 100 and at or about 50 nm, or between at or about 20 nm and at or about 50 nm.

**[0238]** In some embodiments, the lipid particles have an average geometric radius, e.g. as determined by a multi-angle light scattering, of about 100 nm to about two microns. In some embodiments, the lipid particles have an average geometric radius between at or about 2 μm and at or about 1 μm, between at or about 1 μm and at or about 900 nm, between at or about 900 nm and at or about 800 nm, between at or about 800 and at or about 700 nm, between at or about 700 nm and at or about 600 nm, between at or about 600 nm and at or about 500 nm, between at or about 500 nm and at or about 400 nm, between at or about 400 nm and at or about 300 nm, between at or about 300 nm and at or about 200 nm, between at or about 200 and at or about 100 nm, between at or about 100 and at or about 50 nm, or between at or about 20 nm and at or about 50 nm.

**[0239]** In some embodiments, the coefficient of variation (COV) (i.e. standard deviation divided by the mean) of a composition of lipid particles is less than at or about 30%, less than at or about 25%, less than at or about 20%, less than at or about 15%, less than at or about 10% or less than at or about 5%.

**[0240]** In some embodiment, provided compositions of lipid particles are characterized by their polydispersity index, which is a measure of the size distribution of the particles wherein values between 1 (maximum dispersion) and 0 (identical size of all of the particles) are possible. In some embodiments, compositions of lipid particles provided herein have a polydispersity index of between at or about 0.05 and at or about 0.7, between at or about 0.05 and at or about 0.6, between at or about 0.05 and at or about 0.5, between at or about 0.05 and at or about 0.4, between at or about 0.05 and at or about 0.3, between at or about 0.05 and at or about 0.2, between at or about 0.05 and at or about 0.1, between at or about 0.1 and at or about 0.7, between at or about 0.1 and at or about 0.6, between at or about 0.1 and at or about 0.5, between at or about 0.1 and at or about 0.4, between at or about 0.1 and at or about 0.3, between at or about 0.1 and at or about 0.2, between at or about 0.2 and at or about 0.7, between at or about 0.2 and at or about 0.6, between at or about 0.2 and at or about 0.5, between at or about 0.2 and at or about 0.4 between at or about 0.2 and at or about 0.3, between at or about 0.3 and at or about 0.7, between at or about 0.3 and at or about 0.6, between at or about 0.3 and at or about 0.5, between at or about 0.3 and at or about 0.4, between at or about 0.4 and at or about 0.7, between at or about 0.4 and at or about 0.6, between at or about 0.4 and at or about 0.5, between at or about 0.5 and at or about 0.7, between at or about 0.5 and at or about 0.6, or between at or about 0.6 and at or about 0.7. In some embodiments, the polydispersity index is less than at or about 0.05, less than at or about 0.1, less than at or about 0.15, less than at or about 0.2, less than at or about 0.25, less than at or about 0.3, less than at or about 0.4, less than at or about 0.5, less than at or about 0.6 or less than at or about 0.7. Various lipid particles are known, any of which can be generated in accord with the provided embodiments. Non-limiting examples of lipid particles include any as described in, or contain features as described in, International published PCT Application No. WO 2017/095946; WO 2017/095944; WO 2017/095940; WO 2019/157319; WO 2018/208728; WO 2019/113512; WO 2019/161281; WO 2020/102578; WO 2019/222403; WO 2020/014209; WO 2020/102485; WO 2020/102499; WO 2020/102503; WO 2013/148327; WO

2017/182585; WO 2011/058052; or WO 2017/068077, each of which are incorporated by reference in their entirety.

**[0241]** Features of the provided lipid particles are described in the following subsections.

### A. Viral-Based Particles

**[0242]** Provided herein are viral-based particles derived from a virus, including those derived from retroviruses or lentiviruses, containing component of a universal adapter system, such as described in Section I. In some embodiments, the lipid particle's bilayer of amphipathic lipids is or comprises the viral envelope. In some embodiments, the lipid particle's bilayer of amphipathic lipids is or comprises lipids derived from a producer cell. In some embodiments, the viral envelope may comprise a fusogen, e.g., a fusogen that is endogenous to the virus or a pseudotyped fusogen. In some embodiments, the lipid particle's lumen or cavity comprises a viral nucleic acid, e.g., a retroviral nucleic acid, e.g., a lentiviral nucleic acid. In some embodiments, the viral nucleic acid may be a viral genome. In some embodiments, the lipid particle further comprises one or more viral non-structural proteins, e.g., in its cavity or lumen. In some embodiments, the viral-based particle is or comprises a virus-like particle (VLP). In some embodiments, the VLP does not comprise any viral genetic material. In some embodiments, the viral-based particle does not contain any virally derived nucleic acids or viral proteins, such as viral structural proteins.

**[0243]** Biological methods for introducing an exogenous agent to a host cell include the use of DNA and RNA vectors. DNA and RNA vectors can also be used to house and deliver polynucleotides and polypeptides. Viral vectors and virus like particles, and especially retroviral vectors, have become the most widely used method for inserting genes into mammalian, e.g., human cells. Other viral vectors and virus like particles can be derived from lentivirus, poxviruses, herpes simplex virus I, adenoviruses and adeno-associated viruses, and the like. See, for example, U.S. Pat. Nos. 5,350,674 and 5,585,362. Methods for producing cells comprising vectors and/or exogenous acids are well-known in the art. See, for example, Sambrook et al., 2001, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, New York.

**[0244]** In some embodiments, the viral particles or virus-like particles bilayer of amphipathic lipids is or comprises lipids derived from an infected host cell. In some embodiments, the lipid bilayer is a viral envelope. In some embodiments, the viral particles or virus-like particles envelope is obtained from a host cell. In some embodiments, the viral particles or virus-like particles envelope is obtained by the viral capsid from the source cell plasma membrane. In some embodiments, the lipid bilayer is obtained from a membrane other than the plasma membrane of a host cell. In some embodiments, the viral particles or virus-like particles envelope lipid bilayer is embedded with viral proteins, including viral glycoproteins, including a paramyxovirus envelope attachment protein (e.g., NiV-G fused to a component of a universal adapter system).

**[0245]** In some embodiments, one or more transducing units of viral particles or virus-like particles, e.g. retroviral particles or retroviral-like particles, are administered to the subject. In some embodiments, at least 1, 10, 100, 1000, 104, 105, 106, 107, 108, 109, 1010, 1011, 1012, 1013, or 1014, transducing units per kg are administered to the subject. In some embodiments at least 1, 10, 100, 1000, 104, 105, 106, 107, 108, 109, 1010, 1011, 1012, 1013, or 1014, transducing units per target cell per ml of blood are administered to the subject.

### *I. Viral Vector Particles*

**[0246]** In some embodiments, the lipid particle is or comprises a virus or a viral vector, e.g., a retrovirus or retroviral vector, e.g., a lentivirus or lentiviral vector. In some embodiments, the virus or viral vector is recombinant. For instance, the viral particle may be referred to as a recombinant virus and/or a recombinant viral vector, which are used interchangeably. In some embodiments, the lipid particle is a recombinant lentivirus vector particle.

**[0247]** In some embodiments, a lipid particle comprises a lipid bilayer comprising a retroviral vector comprising an envelope. For instance, in some embodiments, the bilayer of amphipathic lipids is or comprises the viral envelope. The viral envelope may comprise a fusogen, e.g., a NiV-G fused to a component of a universal adapter system, that is endogenous to the virus or is a pseudotyped fusogen. In some embodiments, the viral vector's lumen or cavity comprises a viral nucleic acid, e.g., a retroviral nucleic acid, e.g., a lentiviral nucleic acid. The viral nucleic acid may be a viral genome. In some embodiments, the viral vector may further comprises one or more viral non- structural proteins, e.g., in its cavity or lumen. In some embodiments, the virus based vector particles are lentivirus. In some embodiments, the lentiviral vector particle is Human Immunodeficiency Virus-1 (HIV-1).

**[0248]** In some aspects, the viral vector particle is limited in the number of polynucleotides that can be packaged. In some embodiments, nucleotides encoding polypeptides to be packaged can be modified such that they retain functional activity with fewer nucleotides in the coding region than that which encodes for the wild-type peptide. Such modifications can include truncations, or other deletions. In some embodiments, more than one polypeptide can be expressed from the same promoter, such that they are fusion polypeptides. In some embodiments, the insert size to be packaged (i.e., viral genome, or portions thereof; or heterologous polynucleotides as described) can be between 500-1000, 1000-2000, 2000-3000, 3000-4000, 4000-5000, 5000-6000, 6000-7000, or 7000-8000 nucleotides in length. In some embodiments, the insert can be over 8000 nucleotides, such as 9000, 10,000, or 11,000 nucleotides in length.

**[0249]** In some embodiments, the viral vector particle, such as retroviral vector particle, comprises one or more of gag polyprotein, polymerase (e.g., pol), integrase (e.g., a functional or non-functional variant), protease, and a fusogen. In some embodiments, the lipid particle further comprises rev. In some embodiments, one or more of the aforesaid proteins are encoded in the retroviral genome (i.e., the insert

as described above), and in some embodiments, one or more of the aforesaid proteins are provided in trans, e.g., by a helper cell, helper virus, or helper plasmid. In some embodiments, the lipid particle nucleic acid (e.g., retroviral nucleic acid) comprises one or more of the following nucleic acid sequences: 5' LTR (e.g., comprising U5 and lacking a functional U3 domain), Psi packaging element (Psi), Central polypurine tract (cPPT) Promoter operatively linked to the payload gene, payload gene (optionally comprising an intron before the open reading frame), Poly A tail sequence, WPRE, and 3' LTR (e.g., comprising U5 and lacking a functional U3). In some embodiments, the lipid particle nucleic acid further comprises a retroviral cis-acting RNA packaging element, and a cPPT/CTS element. In some embodiments the lipid particle nucleic acid further comprises one or more insulator element. In some embodiments, the recognition sites are situated between the poly A tail sequence and the WPRE.

**[0250]** In some embodiments, the lipid particle comprises supramolecular complexes formed by viral proteins that self-assemble into capsids. In some embodiments, the lipid particle is a viral particle derived from viral capsids. In some embodiments, the lipid particle is a viral particle derived from viral nucleocapsids. In some embodiments, the lipid particle comprises nucleocapsid-derived that retain the property of packaging nucleic acids.

**[0251]** In some embodiments, the lipid particle packages nucleic acids from host cells carrying one or more viral nucleic acids (e.g. retroviral nucleic acids) during the expression process. In some embodiments, the nucleic acids do not encode any genes involved in virus replication. In particular embodiments, the lipid particle is a virus-based particle, e.g. retrovirus particle such as a lentivirus particle, that is replication defective.

**[0252]** In some cases, the lipid particle is a viral particle that is morphologically indistinguishable from the wild type infectious virus. In some embodiments, the viral particle presents the entire viral proteome as an antigen. In some embodiments, the viral particle presents only a portion of the proteome as an antigen.

**[0253]** In some embodiments, the retroviral nucleic acid comprises one or more of (e.g., all of): a 5' promoter (e.g., to control expression of the entire packaged RNA), a 5' LTR (e.g., that includes R (polyadenylation tail signal) and/or U5 which includes a primer activation signal), a primer binding site, a psi packaging signal, a RRE element for nuclear export, a promoter directly upstream of the transgene to control transgene expression, a transgene (or other exogenous agent element), a polypurine tract, and a 3' LTR (e.g., that includes a mutated U3, a R, and U5). In some embodiments, the retroviral nucleic acid further comprises one or more of a cPPT, a WPRE, and/or an insulator element.

**[0254]** A retrovirus typically replicates by reverse transcription of its genomic RNA into a linear double-stranded DNA copy and subsequently covalently integrates its genomic DNA into a host genome. Illustrative retroviruses suitable for use in particular embodiments, include, but are not limited to: Moloney murine leukemia virus (M-MuLV), Moloney murine sarcoma virus (MoMSV), Harvey murine sarcoma virus (HaMuSV), murine mammary tumor virus (MuMTV), gibbon ape leukemia virus (GaLV),

feline leukemia virus (FLV), spumavirus, Friend murine leukemia virus, Murine Stem Cell Virus (MSCV) and Rous Sarcoma Virus (RSV)) and lentivirus.

**[0255]** In some embodiments the retrovirus is a Gammaretrovirus. In some embodiments the retrovirus is an Epsilonretrovirus. In some embodiments the retrovirus is an Alpharetrovirus. In some embodiments the retrovirus is a Betaretrovirus. In some embodiments the retrovirus is a Deltaretrovirus. In some embodiments the retrovirus is a Lentivirus. In some embodiments the retrovirus is a Spumaretrovirus. In some embodiments the retrovirus is an endogenous retrovirus.

**[0256]** Illustrative lentiviruses include, but are not limited to: HIV (human immunodeficiency virus; including HIV type 1, and HIV type 2); visna-maedi virus (VMV) virus; the caprine arthritis-encephalitis virus (CAEV); equine infectious anemia virus (EIAV); feline immunodeficiency virus (FIV); bovine immune deficiency virus (BIV); and simian immunodeficiency virus (SIV). In some embodiments, HIV based vector backbones (i.e., HIV cis-acting sequence elements) are used.

**[0257]** A viral vector can comprise a nucleic acid molecule (e.g., a transfer plasmid) that includes virus-derived nucleic acid elements that typically facilitate transfer of a nucleic acid molecule (e.g. including nucleic acid encoding an exogenous agent) or integration into the genome of a cell or to a viral particle that mediates nucleic acid transfer. Viral vector particles will typically include various viral components and sometimes also host cell components in addition to nucleic acid(s). A viral vector can comprise a virus or viral particle capable of transferring a nucleic acid into a cell (e.g. nucleic acid encoding an exogenous agent), or to the transferred nucleic acid (e.g., as naked DNA). Viral vectors and transfer plasmids can comprise structural and/or functional genetic elements that are primarily derived from a virus. A retroviral vector can comprise a viral vector or plasmid containing structural and functional genetic elements, or portions thereof, that are primarily derived from a retrovirus. A lentiviral vector can comprise a viral vector or plasmid containing structural and functional genetic elements, or portions thereof, including LTRs that are primarily derived from a lentivirus.

**[0258]** In embodiments, a lentiviral vector (e.g., lentiviral expression vector) may comprise a lentiviral transfer plasmid (e.g., as naked DNA) or an infectious lentiviral particle. With respect to elements such as cloning sites, promoters, regulatory elements, heterologous nucleic acids, etc., it is to be understood that the sequences of these elements can be present in RNA form in lentiviral particles and can be present in DNA form in DNA plasmids.

**[0259]** In some vectors described herein, at least part of one or more protein coding regions that contribute to or are essential for replication may be absent compared to the corresponding wild- type virus. This makes the viral vector replication-defective. In some embodiments, the vector is capable of transducing a target non-dividing host cell and/or integrating its genome into a host genome.

**[0260]** The structure of a wild-type retrovirus genome often comprises a 5' long terminal repeat (LTR) and a 3' LTR, between or within which are located a packaging signal to enable the genome to be packaged, a primer binding site, integration sites to enable integration into a host cell genome and gag,

pol and env genes encoding the packaging components which promote the assembly of viral particles. More complex retroviruses have additional features, such as rev and RRE sequences in HIV, which enable the efficient export of RNA transcripts of the integrated provirus from the nucleus to the cytoplasm of an infected target cell. In the provirus, the viral genes are flanked at both ends by regions called long terminal repeats (LTRs). The LTRs are involved in proviral integration and transcription. LTRs also serve as enhancer-promoter sequences and can control the expression of the viral genes. Encapsidation of the retroviral RNAs occurs by virtue of a psi sequence located at the 5' end of the viral genome.

**[0261]** The LTRs themselves are typically similar (e.g., identical) sequences that can be divided into three elements, which are called U3, R and U5. U3 is derived from the sequence unique to the 3' end of the RNA. R is derived from a sequence repeated at both ends of the RNA and U5 is derived from the sequence unique to the 5' end of the RNA. The sizes of the three elements can vary considerably among different retroviruses.

**[0262]** For the viral genome, the site of transcription initiation is typically at the boundary between U3 and R in one LTR and the site of poly (A) addition (termination) is at the boundary between R and U5 in the other LTR. U3 contains most of the transcriptional control elements of the provirus, which include the promoter and multiple enhancer sequences responsive to cellular and in some cases, viral transcriptional activator proteins. Some retroviruses comprise any one or more of the following genes that code for proteins that are involved in the regulation of gene expression: tot, rev, tax and rex. With regard to the structural genes gag, pol and env themselves, gag encodes the internal structural protein of the virus. Gag protein is proteolytically processed into the mature proteins MA (matrix), CA (capsid) and NC (nucleocapsid). The pol gene encodes the reverse transcriptase (RT), which contains DNA polymerase, associated RNase H and integrase (IN), which mediate replication of the genome. The env gene encodes the surface (SU) glycoprotein and the transmembrane (TM) protein of the virion, which form a complex that interacts specifically with cellular receptor proteins. This interaction promotes infection, e.g., by fusion of the viral membrane with the cell membrane.

**[0263]** In a replication-defective retroviral vector genome gag, pol and env may be absent or not functional. The R regions at both ends of the RNA are typically repeated sequences. U5 and U3 represent unique sequences at the 5' and 3' ends of the RNA genome respectively.

**[0264]** Retroviruses may also contain additional genes which code for proteins other than gag, pol and env. Examples of additional genes include (in HIV), one or more of vif, vpr, vpx, vpu, tat, rev and nef. EIAV has (amongst others) the additional gene S2. Proteins encoded by additional genes serve various functions, some of which may be duplicative of a function provided by a cellular protein. In EIAV, for example, tat acts as a transcriptional activator of the viral LTR (Derse and Newbold 1993 Virology 194:530-6; Maury et al. 1994 Virology 200:632- 42). It binds to a stable, stem-loop RNA secondary structure referred to as TAR. Rev regulates and co-ordinates the expression of viral genes

through rev-response elements (RRE) (Martarano et al. 1994 J. Virol. 68:3102-11). The mechanisms of action of these two proteins are thought to be broadly similar to the analogous mechanisms in the primate viruses. In addition, an EIAV protein, Ttm, has been identified that is encoded by the first exon of tat spliced to the env coding sequence at the start of the transmembrane protein.

**[0265]** In addition to protease, reverse transcriptase and integrase, non-primate lentiviruses contain a fourth pol gene product which codes for a dUTPase. This may play a role in the ability of these lentiviruses to infect certain non-dividing or slowly dividing cell types.

**[0266]** In embodiments, a recombinant lentiviral vector (RLV) is a vector with sufficient retroviral genetic information to allow packaging of an RNA genome, in the presence of packaging components, into a viral particle capable of infecting a target cell. Infection of the target cell can comprise reverse transcription and integration into the target cell genome. The RLV typically carries non- viral coding sequences which are to be delivered by the vector to the target cell, such as nucleic acid encoding an exogenous agent as described herein. In embodiments, an RLV is incapable of independent replication to produce infectious retroviral particles within the target cell. Usually the RLV lacks a functional gag-pol and/or env gene and/or other genes involved in replication. The vector may be configured as a split-intron vector, e.g., as described in PCT patent application WO 99/15683, which is herein incorporated by reference in its entirety.

**[0267]** In some embodiments, the lentiviral vector comprises a minimal viral genome, e.g., the viral vector has been manipulated so as to remove the non-essential elements and to retain the essential elements in order to provide the required functionality to infect, transduce and deliver a nucleotide sequence of interest to a target host cell, e.g., as described in WO 98/17815, which is herein incorporated by reference in its entirety.

**[0268]** A minimal lentiviral genome may comprise, e.g., (5')R-U5-one or more first nucleotide sequences-U3-R(3')· However, the plasmid vector used to produce the lentiviral genome within a source cell can also include transcriptional regulatory control sequences operably linked to the lentiviral genome to direct transcription of the genome in a source cell. These regulatory sequences may comprise the natural sequences associated with the transcribed retroviral sequence, e.g., the 5' U3 region, or they may comprise a heterologous promoter such as another viral promoter, for example the CMV promoter. Some lentiviral genomes comprise additional sequences to promote efficient virus production. For example, in the case of HIV, rev and RRE sequences may be included. Alternatively or in combination, codon optimization may be used, e.g., the gene encoding the exogenous agent may be codon optimized, e.g., as described in WO 01/79518, which is herein incorporated by reference in its entirety. Alternative sequences which perform a similar or the same function as the rev/RRE system may also be used. For example, a functional analogue of the rev/RRE system is found in the Mason Pfizer monkey virus. This is known as CTE and comprises an RRE-type sequence in the genome which is believed to interact with a factor in the infected cell. The cellular factor can be thought of as a rev analogue. Thus, CTE may be

used as an alternative to the rev/RRE system. In addition, the Rex protein of HTLV-I can functionally replace the Rev protein of HIV-I . Rev and Rex have similar effects to IRE-BP.

**[0269]** In some embodiments, a retroviral nucleic acid (e.g., a lentiviral nucleic acid, e.g., a primate or non-primate lentiviral nucleic acid) (1) comprises a deleted gag gene wherein the deletion in gag removes one or more nucleotides downstream of about nucleotide 350 or 354 of the gag coding sequence; (2) has one or more accessory genes absent from the retroviral nucleic acid; (3) lacks the tat gene but includes the leader sequence between the end of the 5' LTR and the ATG of gag; and (4) combinations of (1), (2) and (3). In an embodiment the lentiviral vector comprises all of features (1) and (2) and (3). This strategy is described in more detail in WO 99/32646, which is herein incorporated by reference in its entirety.

**[0270]** In some embodiments, a primate lentivirus minimal system requires none of the HIV/SIV additional genes vif, vpr, vpx, vpu, tat, rev and nef for either vector production or for transduction of dividing and non-dividing cells. In some embodiments, an EIAV minimal vector system does not require S2 for either vector production or for transduction of dividing and non dividing cells.

**[0271]** The deletion of additional genes may permit vectors to be produced without the genes associated with disease in lentiviral (e.g. HIV) infections. In particular, tat is associated with disease. Secondly, the deletion of additional genes permits the vector to package more heterologous DNA. Thirdly, genes whose function is unknown, such as S2, may be omitted, thus reducing the risk of causing undesired effects. Examples of minimal lentiviral vectors are disclosed in WO 99/32646 and in WO 98/17815.

**[0272]** In some embodiments, the retroviral nucleic acid is devoid of at least tat and S2 (if it is an EIAV vector system), and possibly also vif, vpr, vpx, vpu and nef. In some embodiments, the retroviral nucleic acid is also devoid of rev, RRE, or both.

**[0273]** In some embodiments the retroviral nucleic acid comprises vpx. The Vpx polypeptide binds to and induces the degradation of the SAMHD1 restriction factor, which degrades free dNTPs in the cytoplasm. Thus, the concentration of free dNTPs in the cytoplasm increases as Vpx degrades SAMHD1 and reverse transcription activity is increased, thus facilitating reverse transcription of the retroviral genome and integration into the target cell genome.

**[0274]** Different cells differ in their usage of particular codons. This codon bias corresponds to a bias in the relative abundance of particular tRNAs in the cell type. By altering the codons in the sequence so that they are tailored to match with the relative abundance of corresponding tRNAs, it is possible to increase expression. By the same token, it is possible to decrease expression by deliberately choosing codons for which the corresponding tRNAs are known to be rare in the particular cell type. Thus, an additional degree of translational control is available. An additional description of codon optimization is found, e.g., in WO 99/41397, which is herein incorporated by reference in its entirety.

**[0275]** Many viruses, including HIV and other lentiviruses, use a large number of rare codons and by changing these to correspond to commonly used mammalian codons, increased expression of the packaging components in mammalian producer cells can be achieved.

**[0276]** In some embodiments, codon optimization has a number of other advantages. In some embodiments, by virtue of alterations in their sequences, the nucleotide sequences encoding the packaging components may have RNA instability sequences (INS) reduced or eliminated from them. At the same time, the amino acid sequence coding sequence for the packaging components is retained so that the viral components encoded by the sequences remain the same, or at least sufficiently similar that the function of the packaging components is not compromised. In some embodiments, codon optimization also overcomes the Rev/RRE requirement for export, rendering optimized sequences Rev independent. In some embodiments, codon optimization also reduces homologous recombination between different constructs within the vector system (for example between the regions of overlap in the gag-pol and env open reading frames). In some embodiments, codon optimization leads to an increase in viral titer and/or improved safety.

**[0277]** In some embodiments, only codons relating to INS are codon optimized. In other embodiments, the sequences are codon optimized in their entirety, with the exception of the sequence encompassing the frameshift site of gag-pol.

**[0278]** The gag-pol gene comprises two overlapping reading frames encoding the gag-pol proteins. The expression of both proteins depends on a frameshift during translation. This frameshift occurs as a result of ribosome "slippage" during translation. This slippage is thought to be caused at least in part by ribosome-stalling RNA secondary structures. Such secondary structures exist downstream of the frameshift site in the gag-pol gene. For HIV, the region of overlap extends from nucleotide 1222 downstream of the beginning of gag (wherein nucleotide 1 is the A of the gag ATG) to the end of gag (nt 1503). Consequently, a 281 bp fragment spanning the frameshift site and the overlapping region of the two reading frames is preferably not codon optimized. In some embodiments, retaining this fragment will enable more efficient expression of the gag-pol proteins. For EIAV, the beginning of the overlap is at nt 1262 (where nucleotide 1 is the A of the gag ATG). The end of the overlap is at nt 1461. In order to ensure that the frameshift site and the gag-pol overlap are preserved, the wild type sequence may be retained from nt 1156 to 1465.

**[0279]** In some embodiments, derivations from optimal codon usage may be made, for example, in order to accommodate convenient restriction sites, and conservative amino acid changes may be introduced into the gag-pol proteins.

**[0280]** In some embodiments, codon optimization is based on codons with poor codon usage in mammalian systems. The third and sometimes the second and third base may be changed.

**[0281]** In some embodiments, due to the degenerate nature of the genetic code, it will be appreciated that numerous gag-pol sequences can be achieved by a skilled worker. Also, there are many retroviral

variants described which can be used as a starting point for generating a codon optimized gag-pol sequence. Lentiviral genomes can be quite variable. For example there are many quasi-species of HIV-I which are still functional. This is also the case for EIAV. These variants may be used to enhance particular parts of the transduction process. Examples of HIV-I variants may be found in the HIV databases maintained by Los Alamos National Laboratory. Details of EIAV clones may be found at the NCBI database maintained by the National Institutes of Health.

**[0282]** It is within the level of a skilled artisan to empirically determine appropriate codon optimization of viral sequences. The strategy for codon optimized sequences, including gag-pol sequences, can be used in relation to any retrovirus, e.g., EIAV, FIV, BIV, CAEV, VMR, SIV, HIV-I and HIV -2. In addition this method could be used to increase expression of genes from HTLV-I, HTLV-2, HFV, HSRV and human endogenous retroviruses (HERV), MLV and other retroviruses.

**[0283]** In embodiments, the retroviral vector comprises a packaging signal that comprises from 255 to 360 nucleotides of gag in vectors that still retain env sequences, or about 40 nucleotides of gag in a particular combination of splice donor mutation, gag and env deletions. In some embodiments, the retroviral vector includes a gag sequence which comprises one or more deletions, e.g., the gag sequence comprises about 360 nucleotides derivable from the N-terminus.

**[0284]** In some embodiments, the retroviral vector, helper cell, helper virus, or helper plasmid may comprise retroviral structural and accessory proteins, for example gag, pol, env, tat, rev, vif, vpr, vpu, vpx, or nef proteins or other retroviral proteins. In some embodiments the retroviral proteins are derived from the same retrovirus. In some embodiments the retroviral proteins are derived from more than one retrovirus, e.g. 2, 3, 4, or more retroviruses.

**[0285]** In some embodiments, the gag and pol coding sequences are generally organized as the Gag-Pol Precursor in native lentivirus. The gag sequence codes for a 55-kD Gag precursor protein, also called p55. The p55 is cleaved by the virally encoded protease (a product of the pol gene) during the process of maturation into four smaller proteins designated MA (matrix [p17]), CA (capsid [p24]), NC (nucleocapsid [p9]), and p6. The pol precursor protein is cleaved away from Gag by a virally encoded protease, and further digested to separate the protease (p10), RT (p50), RNase H (p15), and integrase (p31) activities.

**[0286]** In some embodiments, the lentiviral vector is integration-deficient. In some embodiments, the pol is integrase deficient, such as by encoding due to mutations in the integrase gene. For example, the pol coding sequence can contain an inactivating mutation in the integrase, such as by mutation of one or more of amino acids involved in catalytic activity, i.e. mutation of one or more of aspartic 64, aspartic acid 116 and/or glutamic acid 152. In some embodiments, the integrase mutation is a D64V mutation. In some embodiments, the mutation in the integrase allows for packaging of viral RNA into a lentivirus. In some embodiments, the mutation in the integrase allows for packaging of viral proteins into a lentivirus. In some embodiments, the mutation in the integrase reduces the possibility of insertional mutagenesis. In

some embodiments, the mutation in the integrase decreases the possibility of generating replication-competent recombinants (RCRs) (Wanisch et al. 2009. Mol Ther. 1798):1316-1332).In some embodiments, native Gag-Pol sequences can be utilized in a helper vector (e.g., helper plasmid or helper virus), or modifications can be made. These modifications include, chimeric Gag-Pol, where the Gag and Pol sequences are obtained from different viruses (e.g., different species, subspecies, strains, clades, etc.), and/or where the sequences have been modified to improve transcription and/or translation, and/or reduce recombination.

**[0287]** In some embodiments, the retroviral nucleic acid includes a polynucleotide encoding a 150-250 (e.g., 168) nucleotide portion of a gag protein that (i) includes a mutated INS1 inhibitory sequence that reduces restriction of nuclear export of RNA relative to wild-type INS1, (ii) contains two nucleotide insertion that results in frame shift and premature termination, and/or (iii) does not include INS2, INS3, and INS4 inhibitory sequences of gag.

**[0288]** In some embodiments, a vector described herein is a hybrid vector that comprises both retroviral (e.g., lentiviral) sequences and non-lentiviral viral sequences. In some embodiments, a hybrid vector comprises retroviral e.g., lentiviral, sequences for reverse transcription, replication, integration and/or packaging.

**[0289]** According to certain specific embodiments, most or all of the viral vector backbone sequences are derived from a lentivirus, e.g., HIV-l. However, it is to be understood that many different sources of retroviral and/or lentiviral sequences can be used, or combined and numerous substitutions and alterations in certain of the lentiviral sequences may be accommodated without impairing the ability of a transfer vector to perform the functions described herein. A variety of lentiviral vectors are described in Naldini et ah, (l996a, l996b, and 1998); Zufferey et al., (1997); Dull et al., 1998, U.S. Pat. Nos. 6,013,516; and 5,994,136, many of which may be adapted to produce a retroviral nucleic acid.

**[0290]** At each end of the provirus, long terminal repeats (LTRs) are typically found. An LTR typically comprises a domain located at the ends of retroviral nucleic acid which, in their natural sequence context, are direct repeats and contain U3, R and U5 regions. LTRs generally promote the expression of retroviral genes (e.g., promotion, initiation and polyadenylation of gene transcripts) and viral replication. The LTR can comprise numerous regulatory signals including transcriptional control elements, polyadenylation signals and sequences for replication and integration of the viral genome. The viral LTR is typically divided into three regions called U3, R and U5. The U3 region typically contains the enhancer and promoter elements. The U5 region is typically the sequence between the primer binding site and the R region and can contain the polyadenylation sequence. The R (repeat) region can be flanked by the U3 and U5 regions. The LTR is typically composed of U3, R and U5 regions and can appear at both the 5' and 3' ends of the viral genome. In some embodiments, adjacent to the 5' LTR are sequences for reverse transcription of the genome (the tRNA primer binding site) and for efficient packaging of viral RNA into particles (the Psi site).

**[0291]** In some embodiments, a packaging signal can comprise a sequence located within the retroviral genome which mediate insertion of the viral RNA into the viral capsid or particle, see e.g., Clever et al., 1995. J. of Virology, Vol. 69, No. 4; pp. 2101-2109. Several retroviral vectors use a minimal packaging signal (a psi [Ψ] sequence) for encapsidation of the viral genome.

**[0292]** In various embodiments, retroviral nucleic acids comprise modified 5' LTR and/or 3' LTRs. Either or both of the LTR may comprise one or more modifications including, but not limited to, one or more deletions, insertions, or substitutions. Modifications of the 3' LTR are often made to improve the safety of lentiviral or retroviral systems by rendering viruses replication-defective, e.g., virus that is not capable of complete, effective replication such that infective virions are not produced (e.g., replication-defective lentiviral progeny).

**[0293]** In some embodiments, a vector is a self-inactivating (SIN) vector, e.g., replication- defective vector, e.g., retroviral or lentiviral vector, in which the right (3') LTR enhancer- promoter region, known as the U3 region, has been modified (e.g., by deletion or substitution) to prevent viral transcription beyond the first round of viral replication. In some aspects, provided herein is a replication incompetent (also referred to herein as replication defective) vector particle, that cannot participate in replication in the absence of the packaging cell (i.e., viral vector particles are not produced from the transduced cell). In some aspects, this is because the right (3') LTR U3 region can be used as a template for the left (5') LTR U3 region during viral replication and, thus, absence of the U3 enhancer-promoter inhibits viral replication. In embodiments, the 3' LTR is modified such that the U5 region is removed, altered, or replaced, for example, with an exogenous poly(A) sequence The 3' LTR, the 5' LTR, or both 3' and 5' LTRs, may be modified LTRs. Other modifications to the viral vector, i.e., retroviral or lentiviral vector, to render said vector replication incompetent are known in the art.

**[0294]** In some embodiments, the U3 region of the 5' LTR is replaced with a heterologous promoter to drive transcription of the viral genome during production of viral particles. Examples of heterologous promoters which can be used include, for example, viral simian virus 40 (SV40) (e.g., early or late), cytomegalovirus (CMV) (e.g., immediate early), Moloney murine leukemia virus (MoMLV), Rous sarcoma virus (RSV), and herpes simplex virus (HSV) (thymidine kinase) promoters. In some embodiments, promoters are able to drive high levels of transcription in a Tat- independent manner. In certain embodiments, the heterologous promoter has additional advantages in controlling the manner in which the viral genome is transcribed. For example, the heterologous promoter can be inducible, such that transcription of all or part of the viral genome will occur only when the induction factors are present. Induction factors include, but are not limited to, one or more chemical compounds or the physiological conditions such as temperature or pH, in which the host cells are cultured.

**[0295]** In some embodiments, viral vectors comprise a TAR (trans-activation response) element, e.g., located in the R region of lentiviral (e.g., HIV) LTRs. This element interacts with the lentiviral

trans-activator (tat) genetic element to enhance viral replication. However, this element is not required, e.g., in embodiments wherein the U3 region of the 5' LTR is replaced by a heterologous promoter.

**[0296]** The R region, e.g., the region within retroviral LTRs beginning at the start of the capping group (i.e., the start of transcription) and ending immediately prior to the start of the poly A tract can be flanked by the U3 and U5 regions. The R region plays a role during reverse transcription in the transfer of nascent DNA from one end of the genome to the other.

**[0297]** The retroviral nucleic acid can also comprise a FLAP element, e.g., a nucleic acid whose sequence includes the central polypurine tract and central termination sequences (cPPT and CTS) of a retrovirus, e.g., HIV-l or HIV-2. Suitable FLAP elements are described in U.S. Pat. No. 6,682,907 and in Zennou, et ah, 2000, Cell, 101:173, which are herein incorporated by reference in their entireties. During HIV-l reverse transcription, central initiation of the plus-strand DNA at the central polypurine tract (cPPT) and central termination at the central termination sequence (CTS) can lead to the formation of a three- stranded DNA structure: the HIV-l central DNA flap. In some embodiments, the retroviral or lentiviral vector backbones comprise one or more FLAP elements upstream or downstream of the gene encoding the exogenous agent. For example, in some embodiments a transfer plasmid includes a FLAP element, e.g., a FLAP element derived or isolated from HIV-L.

**[0298]** In embodiments, a retroviral or lentiviral nucleic acid comprises one or more export elements, e.g., a cis-acting post-transcriptional regulatory element which regulates the transport of an RNA transcript from the nucleus to the cytoplasm of a cell. Examples of RNA export elements include, but are not limited to, the human immunodeficiency virus (HIV) rev response element (RRE) (see e.g., Cullen et al., 1991. J. Virol. 65: 1053; and Cullen et al., 1991. Cell 58: 423), and the hepatitis B virus post-transcriptional regulatory element (HPRE), which are herein incorporated by reference in their entireties. Generally, the RNA export element is placed within the 3' UTR of a gene, and can be inserted as one or multiple copies.

**[0299]** In some embodiments, expression of heterologous sequences (e.g. nucleic acid encoding an exogenous agent) in viral vectors is increased by incorporating one or more of, e.g., all of, posttranscriptional regulatory elements, polyadenylation sites, and transcription termination signals into the vectors. A variety of posttranscriptional regulatory elements can increase expression of a heterologous nucleic acid at the protein, e.g., woodchuck hepatitis virus posttranscriptional regulatory element (WPRE; Zufferey et al., 1999, J. Virol., 73:2886); the posttranscriptional regulatory element present in hepatitis B virus (HPRE) (Huang et al., Mol. Cell. Biol., 5:3864); and the like (Liu et al., 1995, Genes Dev., 9:1766), each of which is herein incorporated by reference in its entirety. In some embodiments, a retroviral nucleic acid described herein comprises a posttranscriptional regulatory element such as a WPRE or HPRE.

**[0300]** In some embodiments, a retroviral nucleic acid described herein lacks or does not comprise a posttranscriptional regulatory element such as a WPRE or HPRE.

**[0301]** Elements directing the termination and polyadenylation of the heterologous nucleic acid transcripts may be included, e.g., to increases expression of the exogenous agent. Transcription termination signals may be found downstream of the polyadenylation signal. In some embodiments, vectors comprise a polyadenylation sequence 3' of a polynucleotide encoding the exogenous agent. A polyA site may comprise a DNA sequence which directs both the termination and polyadenylation of the nascent RNA transcript by RNA polymerase II. Polyadenylation sequences can promote mRNA stability by addition of a polyA tail to the 3' end of the coding sequence and thus, contribute to increased translational efficiency. Illustrative examples of polyA signals that can be used in a retroviral nucleic acid, include AATAAA, ATT AAA, AGTAAA, a bovine growth hormone polyA sequence (BGHpA), a rabbit b-globin polyA sequence (rPgpA), or another suitable heterologous or endogenous polyA sequence.

**[0302]** In some embodiments, a retroviral or lentiviral vector further comprises one or more insulator elements, e.g., an insulator element described herein.

**[0303]** In various embodiments, the vectors comprise a promoter operably linked to a polynucleotide encoding an exogenous agent. The vectors may have one or more LTRs, wherein either LTR comprises one or more modifications, such as one or more nucleotide substitutions, additions, or deletions. The vectors may further comprise one of more accessory elements to increase transduction efficiency (e.g., a cPPT/FLAP), viral packaging (e.g., a Psi (Y) packaging signal, RRE), and/or other elements that increase exogenous gene expression (e.g., poly (A) sequences), and may optionally comprise a WPRE or HPRE.

**[0304]** In some embodiments, a lentiviral nucleic acid comprises one or more of, e.g., all of, e.g., from 5' to 3', a promoter (e.g., CMV), an R sequence (e.g., comprising TAR), a U5 sequence (e.g., for integration), a PBS sequence (e.g., for reverse transcription), a DIS sequence (e.g., for genome dimerization), a psi packaging signal, a partial gag sequence, an RRE sequence (e.g., for nuclear export), a cPPT sequence (e.g., for nuclear import), a promoter to drive expression of the exogenous agent, a gene encoding the exogenous agent, a WPRE sequence (e.g., for efficient transgene expression), a PPT sequence (e.g., for reverse transcription), an R sequence (e.g., for polyadenylation and termination), and a U5 signal (e.g., for integration).

### *2. Virus-like Particles*

**[0305]** In some embodiments, the viral-based particles are viral-like lipid particles (VLPs) that are derived from virus. In some embodiments, the viral envelope may comprise a fusogen, e.g., a fusogen that is endogenous to the virus or a pseudotyped fusogen, e.g., a paramyxovirus envelope attachment protein fused to a component of a universal adapter system as described in Section I and/or a paramyxovirus fusion (F) protein (e.g., NiV-F). The VLPs include those derived from retroviruses or lentiviruses. While VLPs mimic native virion structure, they lack the viral genomic information

necessary for independent replication within a host cell. Therefore, in some aspects, VLPs are non-infectious. In particular embodiments, a VLP does not contain a viral genome. In some embodiments, the VLP's bilayer of amphipathic lipids is or comprises the viral envelope. In some embodiments, the lipid particle's bilayer of amphipathic lipids is or comprises lipids derived from a cell. In some embodiments, a VLP contains at least one type of structural protein from a virus. In most cases this protein will form a proteinaceous capsid. In some cases the capsid will also be enveloped in a lipid bilayer originating from the cell from which the assembled VLP has been released (e.g. VLPs comprising a human immunodeficiency virus structural protein such as GAG). In some embodiments, the VLP further comprises a targeting moiety as an envelope protein within the lipid bilayer.

**[0306]** In some embodiments, the vector vehicle particle comprises supramolecular complexes formed by viral proteins that self-assemble into capsids. In some embodiments, the vector vehicle particle is a virus-like particle derived from viral capsid proteins. In some embodiments, the vector vehicle particle is a virus-like particle derived from viral nucleocapsid proteins. In some embodiments, the vector vehicle particle comprises nucleocapsid-derived proteins that retain the property of packaging nucleic acids. In some embodiments, the viral-based particles, such as virus-like particles comprises only viral structural glycoproteins among proteins from the viral genome. In some embodiments, the vector vehicle particle does not contain a viral genome.

**[0307]** In some embodiments, the vector vehicle particle packages nucleic acids from host cells during the expression process, such as a nucleic acid encoding an exogenous agent. In some embodiments, the nucleic acids do not encode any genes involved in virus replication. In particular embodiments, the vector vehicle particle is a virus-like particle, e.g. retrovirus-like particle such as a lentivirus-like particle, that is replication defective.

**[0308]** In some embodiments, the vector vehicle particle is a virus-like particle which comprises a sequence that is devoid of or lacking viral RNA may be the result of removing or eliminating the viral RNA from the sequence. In some embodiments, this may be achieved by using an endogenous packaging signal binding site on gag. In some embodiments, the endogenous packaging signal binding site is on pol. In some embodiments, the RNA which is to be delivered will contain a cognate packaging signal. In some embodiments, a heterologous binding domain (which is heterologous to gag) located on the RNA to be delivered, and a cognate binding site located on gag or pol, can be used to ensure packaging of the RNA to be delivered. In some embodiments, the heterologous sequence could be non-viral or it could be viral, in which case it may be derived from a different virus. In some embodiments, the vector particles could be used to deliver therapeutic RNA, in which case functional integrase and/or reverse transcriptase is not required. In some embodiments, the vector particles could also be used to deliver a therapeutic gene of interest, in which case pol is typically included.

**[0309]** In some embodiments, the VLP comprises supramolecular complexes formed by viral proteins that self-assemble into capsids. In some embodiments, the VLP is derived from viral capsids. In

some embodiments, the VLP is derived from viral nucleocapsids. In some embodiments, the VLP is nucleocapsid-derived and retains the property of packaging nucleic acids. In some embodiments, the VLP includes only viral structural glycoproteins. In some embodiments, the VLP does not contain a viral genome.

### *3. Methods of Generating Viral-based Particles*

**[0310]** Large scale viral particle production is often useful to achieve a desired viral titer. Viral particles can be produced by transfecting a transfer vector into a packaging cell line that comprises viral structural and/or accessory genes, e.g., gag, pol, env, tat, rev, vif, vpr, vpu, vpx, or nef genes or other retroviral genes.

**[0311]** In some embodiments, viral vector particles may be produced in multiple cell culture systems including bacteria, mammalian cell lines, insect cell lines, yeast and plant cells. Methods of producing such viral-based particles can also be found in US Pat No. 10,316,295, which is incorporated herein by reference. Exemplary methods for producing viral vector particles are described.

**[0312]** In some embodiments, elements for the production of a viral vector, i.e., a recombinant viral vector such as a replication incompetent lentiviral vector, are included in a packaging cell line or are present on a packaging vector. In some embodiments, viral vectors can include packaging elements, rev, gag, and pol, delivered to the packaging cells line via one or more packaging vectors.

**[0313]** In embodiments, the packaging vector is an expression vector or viral vector that lacks a packaging signal and comprises a polynucleotide encoding one, two, three, four or more viral structural and/or accessory genes. Typically, the packaging vectors are included in a packaging cell, and are introduced into the cell via transfection, transduction or infection. A retroviral, e.g., lentiviral, transfer vector can be introduced into a packaging cell line, via transfection, transduction or infection, to generate a source cell or cell line. The packaging vectors can be introduced into human cells or cell lines by standard methods including, e.g., calcium phosphate transfection, lipofection or electroporation. In some embodiments, the packaging vectors are introduced into the cells together with a dominant selectable marker, such as neomycin, hygromycin, puromycin, blastocidin, zeocin, thymidine kinase, DHFR, Gln synthetase or ADA, followed by selection in the presence of the appropriate drug and isolation of clones. A selectable marker gene can be linked physically to genes encoding by the packaging vector, e.g., by IRES or self-cleaving viral peptides. In some embodiments, the packaging vector is a packaging plasmid.

**[0314]** Producer cell lines (also called packaging cell lines) include cell lines that do not contain a packaging signal, but do stably or transiently express viral structural proteins and replication enzymes (e.g., gag, pol and env) which can package viral particles. Any suitable cell line can be employed, e.g., mammalian cells, e.g., human cells. Suitable cell lines which can be used include, for example, CHO cells, BHK cells, MDCK cells, C3H 10T1/2 cells, FLY cells, Psi-2 cells, BOSC 23 cells, PA317 cells, WEHI cells, COS cells, BSC 1 cells, BSC 40 cells, BMT 10 cells, VERO cells, W138 cells, MRC5 cells,

A549 cells, HT1080 cells, 293 cells, 293T cells, B-50 cells, 3T3 cells, NIH3T3 cells, HepG2 cells, Saos-2 cells, Huh7 cells, HeLa cells, W163 cells, 211 cells, and 211 A cells. In embodiments, the packaging cells are 293 cells, 293T cells, or A549 cells.

**[0315]** In some embodiments, a producer cell (i.e., a source cell line) includes a cell line which is capable of producing recombinant retroviral particles, comprising a packaging cell line and a transfer vector construct comprising a packaging signal. Methods of preparing viral stock solutions are illustrated by, e.g., Y. Soneoka et al. (1995) Nucl. Acids Res. 23:628-633, and N. R. Landau et al. (1992) J. Virol. 66:5110-5113, which are incorporated herein by reference. Infectious virus particles may be collected from the packaging cells, e.g., by cell lysis, or collection of the supernatant of the cell culture. Optionally, the collected virus particles may be enriched or purified.

**[0316]** In some embodiments, the source cell comprises one or more plasmids coding for viral structural proteins and replication enzymes (e.g., gag, pol and env) which can package viral particles (i.e., a packaging plasmid). In some embodiments, the sequences coding for at least two of the gag, pol, and env precursors are on the same plasmid. In some embodiments, the sequences coding for the gag, pol, and env precursors are on different plasmids. In some embodiments, the sequences coding for the gag, pol, and env precursors have the same expression signal, e.g., promoter. In some embodiments, the sequences coding for the gag, pol, and env precursors have a different expression signal, e.g., different promoters. In some embodiments, expression of the gag, pol, and env precursors is inducible. In some embodiments, the plasmids coding for viral structural proteins and replication enzymes are transfected at the same time or at different times. In some embodiments, the plasmids coding for viral structural proteins and replication enzymes are transfected at the same time or at a different time from the packaging vector.

**[0317]** In some embodiments, the source cell line comprises one or more stably integrated viral structural genes. In some embodiments expression of the stably integrated viral structural genes is inducible.

**[0318]** In some embodiments, expression of the viral structural genes is regulated at the transcriptional level. In some embodiments, expression of the viral structural genes is regulated at the translational level. In some embodiments, expression of the viral structural genes is regulated at the post-translational level.

**[0319]** In some embodiments, expression of the viral structural genes is regulated by a tetracycline (Tet)-dependent system, in which a Tet-regulated transcriptional repressor (Tet-R) binds to DNA sequences included in a promoter and represses transcription by steric hindrance (Yao et al, 1998; Jones et al, 2005). Upon addition of doxycycline (dox), Tet-R is released, allowing transcription. Multiple other suitable transcriptional regulatory promoters, transcription factors, and small molecule inducers are suitable to regulate transcription of viral structural genes.

**[0320]** In some embodiments, the third-generation lentivirus components, human immunodeficiency virus type 1 (HIV) Rev, Gag/Pol, and an envelope under the control of Tet- regulated promoters and coupled with antibiotic resistance cassettes are separately integrated into the source cell genome. In some embodiments the source cell only has one copy of each of Rev, Gag/Pol, and an envelope protein integrated into the genome.

**[0321]** In some embodiments a nucleic acid encoding the exogenous agent (e.g., a retroviral nucleic acid encoding the exogenous agent) is also integrated into the source cell genome. In some embodiments a nucleic acid encoding the exogenous agent is maintained episomally. In some embodiments a nucleic acid encoding the exogenous agent is transfected into the source cell that has stably integrated Rev, Gag/Pol, and an envelope protein in the genome. See, e.g., Milani et al. EMBO Molecular Medicine , 2017, which is herein incorporated by reference in its entirety.

**[0322]** In some embodiments, a retroviral nucleic acid described herein is unable to undergo reverse transcription. Such a nucleic acid, in embodiments, is able to transiently express an exogenous agent. The retrovirus or VLP, may comprise a disabled reverse transcriptase protein, or may not comprise a reverse transcriptase protein. In embodiments, the retroviral nucleic acid comprises a disabled primer binding site (PBS) and/or att site. In embodiments, one or more viral accessory genes, including rev, tat, vif, nef, vpr, vpu, vpx and S2 or functional equivalents thereof, are disabled or absent from the retroviral nucleic acid. In embodiments, one or more accessory genes selected from S2, rev and tat are disabled or absent from the retroviral nucleic acid.

**[0323]** Typically, modern retroviral vector systems include viral genomes bearing cis-acting vector sequences for transcription, reverse-transcription, integration, translation and packaging of viral RNA into the viral particles, and (2) producer cells lines which express the trans-acting retroviral gene sequences (e.g., gag, pol and env) needed for production of virus particles. By separating the cis-and trans-acting vector sequences completely, the virus is unable to maintain replication for more than one cycle of infection. Generation of live virus can be avoided by a number of strategies, e.g., by minimizing the overlap between the cis-and trans-acting sequences to avoid recombination.

**[0324]** A virus-like particle (VLP) which comprises a sequence that is devoid of or lacking viral RNA as described in Section III.A.2 may be the result of removing or eliminating the viral RNA from the sequence. Similar to the viral vector particles disclosed in Section II.A.1, VLPs contain a viral outer envelope made from the host cell (i.e., producer cell or source cell) lipid-bi layer as well as at least one viral structural protein. In some embodiments, a viral structural protein refers to any viral protein or fragment thereof which contributes to the structure of the viral core or capsid.

**[0325]** Generally, for viral vector particles, expression of the gag precursor protein alone mediates vector assembly and release. In some aspects, gag proteins or fragments thereof have been demonstrated to assemble into structures analogous to viral cores. In one embodiment this may be achieved by using an endogenous packaging signal binding site on gag. Alternatively, the endogenous packaging signal

binding site is on pol. In this embodiment, the RNA which is to be delivered will contain a cognate packaging signal. In another embodiment, a heterologous binding domain (which is heterologous to gag) located on the RNA to be delivered, and a cognate binding site located on gag or pol, can be used to ensure packaging of the RNA to be delivered. The heterologous sequence could be non-viral or it could be viral, in which case it may be derived from a different virus. The VLP could be used to deliver therapeutic RNA, in which case functional integrase and/or reverse transcriptase is not required. These VLPs could also be used to deliver a therapeutic gene of interest, in which case pol is typically included.

**[0326]** In an embodiment, gag-pol are altered, and the packaging signal is replaced with a corresponding packaging signal. In this embodiment, the particle can package the RNA with the new packaging signal. The advantage of this approach is that it is possible to package an RNA sequence which is devoid of viral sequence for example, RNAi.

**[0327]** An alternative approach is to rely on over-expression of the RNA to be packaged. In one embodiment the RNA to be packaged is over-expressed in the absence of any RNA containing a packaging signal. This may result in a significant level of therapeutic RNA being packaged, and that this amount is sufficient to transduce a cell and have a biological effect.

**[0328]** In some embodiments, a polynucleotide comprises a nucleotide sequence encoding a viral gag protein or retroviral gag and pol proteins, wherein the gag protein or pol protein comprises a heterologous RNA binding domain capable of recognizing a corresponding sequence in an RNA sequence to facilitate packaging of the RNA sequence into a viral vector particle. In some embodiments, the heterologous RNA binding domain comprises an RNA binding domain derived from a bacteriophage coat protein, a Rev protein, a protein of the U 1 small nuclear ribonucleoprotein particle, a Nova protein, a TF111 A protein, a TIS 11 protein, a trp RNA-binding attenuation protein (TRAP) or a pseudouridine synthase.

**[0329]** In some embodiments, the assembly of a viral based vector particle (i.e., a VLP) is initiated by binding of the core protein to a unique encapsidation sequence within the viral genome (e.g. UTR with stem-loop structure). In some embodiments, the interaction of the core with the encapsidation sequence facilitates oligomerization.

**[0330]** In some embodiments, the source cell for VLP production comprises one or more plasmids coding for viral structural proteins (e.g., gag, pol) which can package viral particles (i.e., a packaging plasmid). In some embodiments, the sequences coding for at least two of the gag and pol precursors are on the same plasmid. In some embodiments, the sequences coding for the gag and pol precursors are on different plasmids. In some embodiments, the sequences coding for the gag and pol precursors have the same expression signal, e.g., promoter. In some embodiments, the sequences coding for the gag and pol precursors have a different expression signal, e.g., different promoters. In some embodiments, expression of the gag and pol precursors is inducible.

**[0331]** In some embodiments, formation of VLPs or any viral-based particle, such as described above in Section II, can be detected by any suitable technique known in the art. Examples of such techniques include, e.g., electron microscopy, dynamic light scattering, selective chromatographic separation and/or density gradient centrifugation.

### *4. Methods of Generating Universal Retargeted Lipid Particle*

**[0332]** Provided herein is a method of making a lipid particle comprising a paramyxovirus envelope attachment protein attached to a targeting moiety via a universal adapter system, said system comprising a tag polypeptide component and a catcher polypeptide component, wherein the paramyxovirus envelope attachment protein is fused to one of the components, and the targeting moiety is fused to the other of the components; and at least one paramyxovirus fusion (F) protein. In some embodiments, the method comprises: (a) providing a cell that comprises any of the provided lipid particles above; (b) culturing the cell under condition that allow for production of the lipid particle, (c) separating, enriching, or purifying the lipid particle from the cell, thereby making the lipid particle; (d) mixing the lipid particle with a targeting moiety fused to the other component of the universal adapter system; and (e) removing unbound targeting moiety from the lipid particles to collect lipid particles with the paramyxovirus envelope attachment protein attached to the targeting moiety.

**[0333]** In some embodiments, the method further comprises concentrating the collected lipid particles

## B. Exogenous Agent

**[0334]** In some embodiments, the lipid particle as described herein or pharmaceutical composition comprising same described contains an exogenous agent. In some embodiments, the lipid particle or pharmaceutical composition comprising same described herein contains a nucleic acid that encodes an exogenous agent. In some embodiments, the lipid particle contains the exogenous agent. In some embodiments, the lipid particle contains a nucleic acid that encodes an exogenous agent. Reference to the coding sequence of the nucleic acid encoding the exogenous agent also is referred to herein as a payload gene. In some embodiments, the exogenous agent or the nucleic acid encoding the exogenous agent are present in the lumen of the lipid particle.

**[0335]** In some embodiments, the exogenous agent is a protein or a nucleic acid (e.g., a DNA, a chromosome (e.g. a human artificial chromosome), an RNA, e.g., an mRNA or miRNA). In some embodiments, the exogenous agent comprises or encodes a membrane protein. In some embodiments, the exogenous agent comprises or encodes a therapeutic agent. In some embodiments, the therapeutic agent is chosen from one or more of a protein, e.g., an enzyme, a transmembrane protein, a receptor, or an antibody; a nucleic acid, e.g., DNA, a chromosome (e.g. a human artificial chromosome), RNA, mRNA, siRNA, or miRNA; or a small molecule.

**[0336]** In some embodiments, the lipid particle or pharmaceutical composition delivers to a target cell at least 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 96%, 97%, 98%, or 99% of the exogenous agent (e.g., an exogenous agent comprising or encoding a therapeutic agent) comprised by the lipid particle. In some embodiments, the lipid particle, e.g., fusosome, that contacts, e.g., fuses, with the target cell(s) delivers to the target cell an average of at least 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 96%, 97%, 98%, or 99% of the exogenous agent (e.g., an exogenous agent comprising or encoding a therapeutic agent) comprised by the lipid particles, e.g., fusosomes, that contact, e.g., fuse, with the target cell(s). In some embodiments, the lipid particle composition delivers to a target tissue at least 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 96%, 97%, 98%, or 99% of the exogenous agent (e.g., an exogenous agent comprising or encoding a therapeutic agent) comprised by the lipid particle compositions.

**[0337]** In some embodiments, the exogenous agent is not expressed naturally in the cell from which the lipid particle is derived. In some embodiments, the exogenous agent is expressed naturally in the cell from which the lipid particle is derived. In some embodiments, the exogenous agent is loaded into the lipid particle via expression in the cell from which the lipid particle is derived (e.g. expression from DNA or mRNA introduced via transfection, transduction, or electroporation). In some embodiments, the exogenous is expressed from DNA integrated into the genome or maintained episomally. In some embodiments, expression of the exogenous agent is constitutive. In some embodiments, expression of the exogenous agent is induced. In some embodiments, expression of the exogenous agent is induced immediately prior to generating the lipid particle. In some embodiments, expression of the exogenous agent is induced at the same time as expression of the fusogen.

**[0338]** In some embodiments, the exogenous agent is loaded into the lipid particle via electroporation into the lipid particle itself or into the cell from which the lipid particle is derived. In some embodiments, the exogenous agent is loaded into the lipid particle via transfection (e.g., of a DNA or mRNA encoding the exogenous agent) into the lipid particle itself or into the cell from which the lipid particle is derived.

**[0339]** In some embodiments, the exogenous agent may include one or more nucleic acid sequences, one or more polypeptides, a combination of nucleic acid sequences and/or polypeptides, one or more organelles, and any combination thereof. In some embodiments, the exogenous agent may include one or more cellular components. In some embodiments, the exogenous agent includes one or more cytosolic and/or nuclear components.

**[0340]** In some embodiments, the lipid particle contains an exogenous agent that is a nucleic acid or contains a nucleic acid encoding the exogenous agent. In some embodiments, the nucleic acid is operatively linked to a "positive target cell-specific regulatory element" (or positive TCSRE). In some embodiments, the positive TCSRE is a functional nucleic acid sequence. In some embodiments, the positive TCSRE comprises a promoter or enhancer. In some embodiments, the TCSRE is a nucleic acid

sequence that increases the level of an exogenous agent in a target cell. In some embodiments, the positive target cell-specific regulatory element comprises a T cell-specific promoter, a T cell-specific enhancer, a T cell-specific splice site, a T cell-specific site extending half-life of an RNA or protein, a T cell-specific mRNA nuclear export promoting site, a T cell-specific translational enhancing site, or a T cell-specific post-translational modification site. In some embodiments, the T cell-specific promoter is a promoter described in Immgen consortium, herein incorporated by reference in its entirety, e.g., the T cell-specific promoter is an IL2RA (CD25), LRRC32, FOXP3, or IKZF2 promoter. In some embodiments, the T cell-specific promoter or enhancer is a promoter or enhancer described in Schmidl et a , Blood. 2014 Apr 24;123(17):e68-78., herein incorporated by reference in its entirety. In some embodiments, the T cell-specific promoter is a transcriptionally active fragment of any of the foregoing. In some embodiments, the T-cell specific promoter is a variant having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity to any of the foregoing.

**[0341]** In some embodiments, the lipid particle contains an exogenous agent that is a nucleic acid or contains a nucleic acid encoding the exogenous agent. In some embodiments, the nucleic acid is operatively linked to a "negative target cell-specific regulatory element" (or negative TCSRE). In some embodiments, the negative TCSRE is a functional nucleic acid sequence. In some embodiments, the negative TCSRE is a miRNA recognition site that causes degradation of inhibition of the lipid particle in a non-target cell. In some embodiments, the exogenous agent is operatively linked to a "non-target cell-specific regulatory element" (or NTCSRE). In some embodiments, the NTCSRE comprises a nucleic acid sequence that decreases the level of an exogenous agent in a non-target cell compared to in a target cell. In some embodiments, the NTCSRE comprises a non-target cell-specific miRNA recognition sequence, non-target cell-specific protease recognition site, non-target cell-specific ubiquitin ligase site, non-target cell-specific transcriptional repression site, or non-target cell-specific epigenetic repression site. In some embodiments, the NTCSRE comprises a tissue-specific miRNA recognition sequence, tissue-specific protease recognition site, tissue-specific ubiquitin ligase site, tissue-specific transcriptional repression site, or tissue-specific epigenetic repression site. In some embodiments, the NTCSRE comprises a non-target cell-specific miRNA recognition sequence, non-target cell-specific protease recognition site, non-target cell-specific ubiquitin ligase site, non-target cell-specific transcriptional repression site, or non-target cell-specific epigenetic repression site. In some embodiments, the NTCSRE comprises a non-target cell-specific miRNA recognition sequence and the miRNA recognition sequence is able to be bound by one or more of miR3 1, miR363, or miR29c. In some embodiments, the NTCSRE is situated or encoded within a transcribed region encoding the exogenous agent, optionally wherein an RNA produced by the transcribed region comprises the miRNA recognition sequence within a UTR or coding region.

### *I. Nucleic Acids*

**[0342]** In some embodiments, the exogenous agent may include a nucleic acid. For example, the exogenous agent may comprise RNA to enhance expression of an endogenous protein, or a siRNA or miRNA that inhibits protein expression of an endogenous protein. For example, the endogenous protein may modulate structure or function in the target cells. In some embodiments, the exogenous agent may include a nucleic acid encoding an engineered protein that modulates structure or function in the target cells. In some embodiments, the exogenous agent is a nucleic acid that targets a transcriptional activator that modulate structure or function in the target cells

**[0343]** In some embodiments, a lipid particle described herein comprises a nucleic acid, e.g., RNA or DNA. In some embodiments, the nucleic acid is, comprises, or consists of one or more natural nucleic acid residues. In some embodiments, the nucleic acid is, comprises, or consists of one or more nucleic acid analogs. In some embodiments, the nucleic acid has a nucleotide sequence that encodes a functional gene product such as an RNA or protein. In some embodiments, the nucleic acid includes one or more introns. In some embodiments, nucleic acids are prepared by one or more of isolation from a natural source, enzymatic synthesis by polymerization based on a complementary template (in vivo or in vitro), reproduction in a recombinant cell or system, and chemical synthesis. In some embodiments, the nucleic acid is at least 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 1 10, 120, 130, 140, 150, 160, 170, 180, 190, 20, 225, 250, 275, 300, 325, 350, 375, 400, 425, 450, 475, 500, 600, 700, 800, 900, 1000, 1500, 2000, 2500, 3000, 3500, 4000, 4500, 5000 or more residues long. In some embodiments, the nucleic acid is partly or wholly single stranded; in some embodiments, the nucleic acid is partly or wholly double stranded. In some embodiments the nucleic acid has a nucleotide sequence comprising at least one element that encodes, or is the complement of a sequence that encodes, a polypeptide. The nucleic acid may include variants, e.g., having an overall sequence identity with a reference nucleic acid of at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, or 99%. In some embodiments, a variant nucleic acid does not share at least one characteristic sequence element with a reference nucleic acid. In some embodiments, a variant nucleic acid shares one or more of the biological activities of the reference nucleic acid. In some embodiments, a nucleic acid variant has a nucleic acid sequence that is identical to that of the reference but for a small number of sequence alterations at particular positions. In some embodiments, fewer than about 20%, about 15%, about 10%, about 9%, about 8%, about 7%, about 6%, about 5%, about 4%, about 3%, or about 2% of the residues in a variant are substituted, inserted, or deleted, as compared to the reference. In some embodiments, a variant nucleic acid comprises about 10, about 9, about 8, about 7, about 6, about 5, about 4, about 3, about 2, or about 1 substituted residue as compared to a reference. In some embodiments, a variant nucleic acid comprises a very small number (e.g., fewer than about 5, about 4, about 3, about 2, or about 1) number of substituted, inserted, or deleted, functional residues that

participate in a particular biological activity relative to the reference. In some embodiments, a variant nucleic acid comprises not more than about 15, about 12, about 9, about 3, or about 1 addition or deletion, and, in some embodiments, comprises no additions or deletions, as compared to the reference. In some embodiments, a variant nucleic acid comprises fewer than about 27, about 24, about 21, about 18, about 15, about 12, about 9, about 6, about 3, or fewer than about 9, about 6, about 3, or about 2 additions or deletions as compared to the reference.

**[0344]** In some embodiments, the exogenous agent includes a nucleic acid, e.g., DNA, nDNA (nuclear DNA), mtDNA (mitochondrial DNA), protein coding DNA, gene, operon, chromosome, genome, transposon, retrotransposon, viral genome, intron, exon, modified DNA, mRNA (messenger RNA), tRNA (transfer RNA), modified RNA, microRNA, siRNA (small interfering RNA), tmRNA (transfer messenger RNA), rRNA (ribosomal RNA), mtRNA (mitochondrial RNA), snRNA (small nuclear RNA), small nucleolar RNA (snoRNA), SmY RNA (mRNA trans-splicing RNA), gRNA (guide RNA), TERC (telomerase RNA component), aRNA (antisense RNA), cis-NAT (Cis-natural antisense transcript), CRISPR RNA (crRNA), IncRNA (long noncoding RNA), piRNA (piwi-interacting RNA), shRNA (short hairpin RNA), tasiRNA (trans-acting siRNA), eRNA (enhancer RNA), satellite RNA, pcRNA (protein coding RNA), dsRNA (double stranded RNA), RNAi (interfering RNA), circRNA (circular RNA), reprograming RNAs, aptamers, and any combination thereof. In some embodiments, the nucleic acid is a wild-type nucleic acid. In some embodiments, the protein is a mutant nucleic acid. In some embodiments the nucleic acid is a fusion or chimera of multiple nucleic acid sequences

**[0345]** In embodiments, the nucleic acid encodes one or more (e.g. two or more) inhibitory RNA molecules directed against one or more RNA targets. An inhibitory RNA molecule can be, e.g., a miRNA or an shRNA. In some embodiments, the inhibitory molecule can be a precursor of a miRNA, such as for example, a Pri-miRNA or a Pre-miRNA, or a precursor of an shRNA. In some embodiments, the inhibitory molecule can be an artificially derived miRNA or shRNA. In other embodiments, the inhibitory RNA molecule can be a dsRNA (either transcribed or artificially introduced) that is processed into an siRNA or the siRNA itself. In some embodiments, the inhibitory RNA molecule can be a miRNA or shRNA that has a sequence that is not found in nature, or has at least one functional segment that is not found in nature, or has a combination of functional segments that are not found in nature. In illustrative embodiments, at least one or all of the inhibitory RNA molecules are miR-l55. In some embodiments, a retroviral vector described herein encodes two or more inhibitory RNA molecules directed against one or more RNA targets. Two or more inhibitory RNA molecules, in some embodiments, can be directed against different targets. In other embodiments, the two or more inhibitory RNA molecules are directed against the same target. In some embodiments, the exogenous agent comprises a shRNA. A shRNA (short hairpin RNA) can comprise a double-stranded structure that is formed by a single self complementary RNA strand. shRNA constructs can comprise a nucleotide sequence identical to a portion, of either coding or non-coding sequence, of a target gene. RNA sequences with insertions,

deletions, and single point mutations relative to the target sequence can also be used. Greater than 90% sequence identity, or even 100% sequence identity, between the inhibitory RNA and the portion of the target gene can be used. In certain embodiments, the length of the duplex-forming portion of an shRNA is at least 20, 2 1 or 22 nucleotides in length, e.g., corresponding in size to RNA products produced by Dicer-dependent cleavage. In certain embodiments, the shRNA construct is at least 25, 50, 100, 200, 300 or 400 bases in length. In certain embodiments, the shRNA construct is 400-800 bases in length. shRNA constructs are highly tolerant of variation in loop sequence and loop size. In embodiments, a retroviral vector that encodes an siRNA, an miRNA, an shRNA, or a ribozyme comprises one or more regulatory sequences, such as, for example, a strong constitutive pol III, e.g., human U6 snRNA promoter, the mouse U6 snRNA promoter, the human and mouse H l RNA promoter and the human tRNA-val promoter, or a strong constitutive pol II promoter.

### *2. Polypeptides*

**[0346]** In some embodiments, the lipid particle contains a nucleic acid that encodes a protein exogenous agent (also referred to as a "payload gene encoding an exogenous agent."). In some embodiments, a lipid particle described herein comprises an exogenous agent which is or comprises a protein.

**[0347]** In some embodiments, the protein may include moieties other than amino acids (e.g., may be glycoproteins, proteoglycans, etc.) and/or may be otherwise processed or modified. In some embodiments, the protein can sometimes include more than one polypeptide chain, for example linked by one or more disulfide bonds or associated by other means.

**[0348]** In some embodiments, the protein may contain L-amino acids, D-amino acids, or both and may contain any of a variety of amino acid modifications or analogs. In some embodiments, proteins may comprise natural amino acids, non-natural amino acids, synthetic amino acids, and combinations thereof. In some embodiments, proteins are antibodies, antibody fragments, biologically active portions thereof, and/or characteristic portions thereof. In some embodiments, a polypeptide may include its variants, e.g., having an overall sequence identity with a reference polypeptide of at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, or 99%. In some embodiments, a variant polypeptide does not share at least one characteristic sequence element with a reference polypeptide. In some embodiments, a variant polypeptide shares one or more of the biological activities of the reference polypeptide. In some embodiments, a polypeptide variant has an amino acid sequence that is identical to that of the reference but for a small number of sequence alterations at particular positions. In some embodiments, fewer than about 20%, about 15%, about 10%, about 9%, about 8%, about 7%, about 6%, about 5%, about 4%, about 3%, or about 2% of the residues in a variant are substituted, inserted, or deleted, as compared to the reference. In some embodiments, a variant polypeptide comprises about 10, about 9, about 8, about 7, about 6, about 5, about 4, about 3, about 2, or about 1 substituted residue as

compared to a reference. In some embodiments, a variant polypeptide comprises a very small number (e.g., fewer than about 5, about 4, about 3, about 2, or about 1) number of substituted, inserted, or deleted, functional that participate in a particular biological activity relative to the reference. In some embodiments, a variant polypeptide comprises not more than about 5, about 4, about 3, about 2, or about 1 addition or deletion, and, in some embodiments, comprises no additions or deletions, as compared to the reference. In some embodiments, a variant polypeptide comprises fewer than about 25, about 20, about 19, about 18, about 17, about 16, about 15, about 14, about 13, about 10, about 9, about 8, about 7, about 6, and commonly fewer than about 5, about 4, about 3, or about 2 additions or deletions as compared to the reference. In some embodiments, the protein includes a polypeptide, e.g., enzymes, structural polypeptides, signaling polypeptides, regulatory polypeptides, transport polypeptides, sensory polypeptides, motor polypeptides, defense polypeptides, storage polypeptides, transcription factors, antibodies, cytokines, hormones, catabolic polypeptides, anabolic polypeptides, proteolytic polypeptides, metabolic polypeptides, kinases, transferases, hydrolases, lyases, isomerases, ligases, enzyme modulator polypeptides, protein binding polypeptides, lipid binding polypeptides, membrane fusion polypeptides, cell differentiation polypeptides, epigenetic polypeptides, cell death polypeptides, nuclear transport polypeptides, nucleic acid binding polypeptides, reprogramming polypeptides, DNA editing polypeptides, DNA repair polypeptides, DNA recombination polypeptides, transposase polypeptides, DNA integration polypeptides, targeted endonucleases (e.g. Zinc -finger nucleases, transcription-activator-like nucleases (TALENs), cas9 and homologs thereof), recombinases, and any combination thereof. In some embodiments, the protein targets a protein in the cell for degradation. In some embodiments, the protein targets a protein in the cell for degradation by localizing the protein to the proteasome. In some embodiments, the protein is a wild-type protein. In some embodiments, the protein is a mutant protein.

**[0349]** Exemplary protein exogenous agents are described in the following subsections. In some embodiments, a lipid particle provided herein can include any of such exogenous agents. In particular embodiments, a lipid particle contains a nucleic acid encoding any of such exogenous agents.

a. Cytosolic Proteins

**[0350]** In some embodiments, the exogenous agent comprises a cytosolic protein, e.g., a protein that is produced in the recipient cell and localizes to the recipient cell cytoplasm. In some embodiments, the exogenous agent comprises a secreted protein, e.g., a protein that is produced and secreted by the recipient cell. In some embodiments, the exogenous agent comprises a nuclear protein, e.g., a protein that is produced in the recipient cell and is imported to the nucleus of the recipient cell. In some embodiments, the exogenous agent comprises an organellar protein (e.g., a mitochondrial protein), e.g., a protein that is produced in the recipient cell and is imported into an organelle (e.g., a mitochondrial) of

the recipient cell. In some embodiments, the protein is a wild-type protein or a mutant protein. In some embodiments the protein is a fusion or chimeric protein.

b. Membrane Proteins

**[0351]** In some embodiments, the exogenous agent comprises a membrane protein. In some embodiments, the membrane protein comprises a chimeric antigen receptor (CAR), a T cell receptor, an integrin, an ion channel, a pore forming protein, a Toll-Like Receptor, an interleukin receptor, a cell adhesion protein, or a transport protein.

1) Chimeric Antigen Receptors (CARs)

**[0352]** In some embodiments, a payload gene described herein encodes a chimeric antigen receptor (CAR) comprising an antigen binding domain. In some embodiments, an exogenous agent described herein comprises a chimeric antigen receptor (CAR) comprising an antigen binding domain. In some embodiments, the payload is or comprises a chimeric antigen receptor (CAR) comprising an antigen binding domain. In some embodiments, the CAR is or comprises a first generation CAR comprising an antigen binding domain, a transmembrane domain, and signaling domain (e.g., one, two or three signaling domains). In some embodiments, the CAR comprises a third generation CAR comprising an antigen binding domain, a transmembrane domain, and at least three signaling domains. In some embodiments, a fourth generation CAR comprising an antigen binding domain, a transmembrane domain, three or four signaling domains, and a domain which upon successful signaling of the CAR induces expression of a cytokine gene. In some embodiments, the antigen binding domain is or comprises an scFv or Fab.

**[0353]** In some embodiments, the antigen binding domain targets an antigen characteristic of a cell type. In some embodiments, the antigen binding domain targets an antigen characteristic of a neoplastic cell. In some embodiments, the antigen characteristic of a neoplastic cell is selected from a cell surface receptor, an ion channel-linked receptor, an enzyme-linked receptor, a G protein-coupled receptor, receptor tyrosine kinase, tyrosine kinase associated receptor, receptor-like tyrosine phosphatase, receptor serine/ threonine kinase, receptor guanylyl cyclase, histidine kinase associated receptor, Epidermal Growth Factor Receptors (EGFR) (including ErbB1/EGFR, ErbB2/HER2, ErbB3/HER3, and ErbB4/HER4), Fibroblast Growth Factor Receptors (FGFR) (including FGF1, FGF2, FGF3, FGF4, FGF5, FGF6, FGF7, FGF18, and FGF21) Vascular Endothelial Growth Factor Receptors (VEGFR) (including VEGF-A, VEGF-B, VEGF-C, VEGF-D, and PIGF), RET Receptor and the Eph Receptor Family (including EphA1, EphA2, EphA3, EphA4, EphA5, EphA6, EphA7, EphA8, EphA9, EphA10, EphB1, EphB2. EphB3, EphB4, and EphB6), CXCR1, CXCR2, CXCR3, CXCR4, CXCR6, CCR1, CCR2, CCR3, CCR4, CCR5, CCR6, CCR8, CFTR, CIC-1, CIC-2, CIC-4, CIC-5, CIC-7, CIC-Ka, CIC-Kb, Bestrophins, TMEM16A, GABA receptor, glycin receptor, ABC transporters, NAV1.1, NAV1.2,

NAV1.3, NAV1.4, NAV1.5, NAV1.6, NAV1.7, NAV1.8, NAV1.9, sphingosin-1-phosphate receptor (S1P1R), NMDA channel, transmembrane protein, multispan transmembrane protein, T-cell receptor motifs; T-cell alpha chains; T-cell β chains; T-cell γ chains; T-cell δ chains; CCR7; CD3; CD4; CD5; CD7; CD8; CD11b; CD11c; CD16; CD19; CD20; CD21 ; CD22; CD25; CD28; CD34; CD35; CD40; CD45RA; CD45RO; CD52; CD56; CD62L; CD68; CD80; CD95; CD117; CD127; CD133; CD137 (4-1 BB); CD163; F4/80; IL-4Ra; Sca-1 ; CTLA-4; GITR; GARP; LAP; granzyme B; LFA-1 ; transferrin receptor; NKp46, perforin, CD4+; Th1; Th2; Th17; Th40; Th22; Th9; Tfh, Canonical Treg. FoxP3+; Tr1; Th3; Treg17; $T_{RE}G$; CDCP1, NT5E, EpCAM, CEA, gpA33, Mucins, TAG-72, Carbonic anhydrase IX, PSMA, Folate binding protein, Gangliosides (e.g., CD2, CD3, GM2), Lewis-$\gamma^2$, VEGF, VEGFR 1/2/3, αVβ3, α5β1, ErbB1/EGFR, ErbB1/HER2, ErB3, c-MET, IGF1R, EphA3, TRAIL-R1, TRAIL-R2, RANKL, FAP, Tenascin, PDL-1, BAFF, HDAC, ABL, FLT3, KIT, MET, RET, IL-1β, ALK, RANKL, mTOR, CTLA-4, IL-6, IL-6R, JAK3, BRAF, PTCH, Smoothened, PIGF, ANPEP, TIMP1, PLAUR, PTPRJ, LTBR, or ANTXR1, Folate receptor alpha (FRa), ERBB2 (Her2/neu), EphA2, IL-13Ra2, epidermal growth factor receptor (EGFR), Mesothelin, TSHR, CD19, CD123, CD22, CD30, CD171, CS-1, CLL-1, CD33, EGFRvIII , GD2, GD3, BCMA, MUC16 (CA125), L1CAM, LeY, MSLN, IL13Rα1, L1-CAM, Tn Ag, prostate specific membrane antigen (PSMA), ROR1, FLT3, FAP, TAG72, CD38, CD44v6, CEA, EPCAM, B7H3, KIT, interleukin-11 receptor a (IL-11Ra), PSCA, PRSS21, VEGFR2, LewisY, CD24, platelet-derived growth factor receptor-beta (PDGFR-beta), SSEA-4, CD20, MUC1, NCAM, Prostase, PAP, ELF2M, Ephrin B2, IGF-1 receptor, CAIX, LMP2, gplOO, bcr-abl, tyrosinase, Fucosyl GM1, sLe, GM3, TGS5, HMWMAA, o-acetyl-GD2, Folate receptor beta, TEM1/CD248, TEM7R, CLDN6, GPRC5D, CXORF61, CD97, CD179a, ALK, Polysialic acid, PLACl, GloboH, NY-BR-1, UPK2, HAVCR1, ADRB3, PANX3, GPR20, LY6K, OR51E2, TARP, WT1, NY-ESO-1, LAGE-la, MAGE-A1, legumain, HPV E6, E7, ETV6-AML, sperm protein 17, XAGE1, Tie 2, MAD-CT-1, MAD-CT-2, Major histocompatibility complex class I-related gene protein (MR1), urokinase-type plasminogen activator receptor (uPAR), Fos-related antigen 1, p53, p53 mutant, prostein, survivin, telomerase, PCTA-1/Galectin 8, MelanA/MART1, Ras mutant, hTERT, sarcoma translocation breakpoints, ML-IAP, ERG (TMPRSS2 ETS fusion gene), NA17, PAX3, Androgen receptor, Cyclin B1, MYCN, RhoC, TRP-2, CYPIB I, BORIS, SART3, PAX5, OY-TES1, LCK, AKAP-4, SSX2, RAGE-1, human telomerase reverse transcriptase, RU1, RU2, intestinal carboxyl esterase, mut hsp70-2, CD79a, CD79b, CD72, LAIR1, FCAR, LILRA2, CD300LF, CLEC12A, BST2, EMR2, LY75, GPC3, FCRL5, IGLL1, a neoantigen, CD133, CD15, CD184, CD24, CD56, CD26, CD29, CD44, HLA-A, HLA-B, HLA-C, (HLA-A,B,C) CD49f, CD151 CD340, CD200, tkrA, trkB, or trkC, or an antigenic fragment or antigenic portion thereof.

**[0354]** In some embodiments, the antigen binding domain targets an antigen characteristic of a T cell. In some embodiments, the antigen characteristic of a T cell is selected from a cell surface receptor, a membrane transport protein (e.g., an active or passive transport protein such as, for example, an ion

channel protein, a pore-forming protein, etc.), a transmembrane receptor, a membrane enzyme, and/or a cell adhesion protein characteristic of a T cell. In some embodiments, an antigen characteristic of a T cell may be a G protein-coupled receptor, receptor tyrosine kinase, tyrosine kinase associated receptor, receptor-like tyrosine phosphatase, receptor serine/ threonine kinase, receptor guanylyl cyclase, histidine kinase associated receptor, AKT1; AKT2; AKT3; ATF2; BCL10; CALM1; CD3D (CD3δ); CD3E (CD3ε); CD3G (CD3γ); CD4; CD8; CD28; CD45; CD80 (B7-1); CD86 (B7-2); CD247 (CD3ζ); CTLA4 (CD152); ELK1; ERK1 (MAPK3); ERK2; FOS; FYN; GRAP2 (GADS); GRB2; HLA-DRA; HLA-DRB1; HLA-DRB3; HLA-DRB4; HLA-DRB5; HRAS; IKBKA (CHUK); IKBKB; IKBKE; IKBKG (NEMO); IL2; ITPR1; ITK; JUN; KRAS2; LAT; LCK; MAP2K1 (MEK1); MAP2K2 (MEK2); MAP2K3 (MKK3); MAP2K4 (MKK4); MAP2K6 (MKK6); MAP2K7 (MKK7); MAP3K1 (MEKK1); MAP3K3; MAP3K4; MAP3K5; MAP3K8; MAP3K14 (NIK); MAPK8 (JNK1); MAPK9 (JNK2); MAPK10 (JNK3); MAPK11 (p38β); MAPK12 (p38γ); MAPK13 (p38δ); MAPK14 (p38α); NCK; NFAT1; NFAT2; NFKB1; NFKB2; NFKBIA; NRAS; PAK1; PAK2; PAK3; PAK4; PIK3C2B; PIK3C3 (VPS34); PIK3CA; PIK3CB; PIK3CD; PIK3R1; PKCA; PKCB; PKCM; PKCQ; PLCY1; PRF1 (Perforin); PTEN; RAC1; RAF1; RELA; SDF1; SHP2; SLP76; SOS; SRC; TBK1; TCRA; TEC; TRAF6; VAV1; VAV2; or ZAP70.

**[0355]** In some embodiments, the antigen binding domain targets an antigen characteristic of a disorder. In some embodiments, the antigen binding domain targets an antigen characteristic of an autoimmune or inflammatory disorder. In some embodiments, the autoimmune or inflammatory disorder is selected from chronic graft-vs-host disease (GVHD), lupus, arthritis, immune complex glomerulonephritis, goodpasture, uveitis, hepatitis, systemic sclerosis or scleroderma, type I diabetes, multiple sclerosis, cold agglutinin disease, Pemphigus vulgaris, Grave's disease, autoimmune hemolytic anemia, Hemophilia A, Primary Sjogren's Syndrome, thrombotic thrombocytopenia purrpura, neuromyelits optica, Evan's syndrome, IgM mediated neuropathy, cyroglobulinemia, dermatomyositis, idiopathic thrombocytopenia, ankylosing spondylitis, bullous pemphigoid, acquired angioedema, chronic urticarial, antiphospholipid demyelinating polyneuropathy, and autoimmune thrombocytopenia or neutropenia or pure red cell aplasias, while exemplary non-limiting examples of alloimmune diseases include allosensitization (see, for example, Blazar et al., 2015, Am. J. Transplant, 15(4):931-41) or xenosensitization from hematopoietic or solid organ transplantation, blood transfusions, pregnancy with fetal allosensitization, neonatal alloimmune thrombocytopenia, hemolytic disease of the newborn, sensitization to foreign antigens such as can occur with replacement of inherited or acquired deficiency disorders treated with enzyme or protein replacement therapy, blood products, and gene therapy. In some embodiments, the antigen characteristic of an autoimmune or inflammatory disorder is selected from a cell surface receptor, an ion channel-linked receptor, an enzyme-linked receptor, a G protein-coupled receptor, receptor tyrosine kinase, tyrosine kinase associated receptor, receptor-like tyrosine phosphatase, receptor serine/ threonine kinase, receptor guanylyl cyclase, or histidine kinase associated receptor. In

some embodiments, a CAR antigen binding domain binds to a ligand expressed on B cells, plasma cells, plasmablasts, CD10, CD19, CD20, CD22, CD24, CD27, CD38, CD45R, CD138, CD319, BCMA, CD28, TNF, interferon receptors, GM-CSF, ZAP-70, LFA-1, CD3 gamma, CD5 or CD2. See US 2003/0077249; WO 2017/058753; WO 2017/058850, the contents of which are herein incorporated by reference.

**[0356]** In some embodiments, the antigen binding domain targets an antigen characteristic of senescent cells, e.g., urokinase-type plasminogen activator receptor (uPAR). In some embodiments, the CAR may be used for treatment or prophylaxis of disorders characterized by the aberrant accumulation of senescent cells, e.g., liver and lung fibrosis, atherosclerosis, diabetes and osteoarthritis.

**[0357]** In some embodiments, the antigen binding domain targets an antigen characteristic of an infectious disease. In some embodiments, wherein the infectious disease is selected from HIV, hepatitis B virus, hepatitis C virus, Human herpes virus, Human herpes virus 8 (HHV-8, Kaposi sarcoma-associated herpes virus (KSHV)), Human T-lymphotrophic virus-1 (HTLV-1), Merkel cell polyomavirus (MCV), Simian virus 40 (SV40), Epstein-Barr virus, CMV, human papillomavirus. In some embodiments, the antigen characteristic of an infectious disease is selected from a cell surface receptor, an ion channel-linked receptor, an enzyme-linked receptor, a G protein-coupled receptor, receptor tyrosine kinase, tyrosine kinase associated receptor, receptor-like tyrosine phosphatase, receptor serine/ threonine kinase, receptor guanylyl cyclase, histidine kinase associated receptor, HIV Env, gpl20, or CD4-induced epitope on HIV-1 Env.

**[0358]** In some embodiments, the CAR transmembrane domain comprises at least a transmembrane region of the alpha, beta or zeta chain of a T cell receptor, CD28, CD3 epsilon, CD45, CD4, CD5, CD8, CD9, CD16, CD22, CD33, CD37, CD64, CD80, CD86, CD134, CD137, CD154, or functional variant thereof. In some embodiments, the transmembrane domain comprises at least a transmembrane region(s) of CD8α, CD8β, 4-1BB/CD137, CD28, CD34, CD4, FcεRIγ, CD16, OX40/CD134, CD3ζ, CD3ε, CD3γ, CD3δ, TCRα, TCRβ, TCRζ, CD32, CD64, CD64, CD45, CD5, CD9, CD22, CD37, CD80, CD86, CD40, CD40L/CD154, VEGFR2, FAS, and FGFR2B, or functional variant thereof.

**[0359]** In some embodiments, the CAR comprises at least one signaling domain selected from one or more of B7-1/CD80; B7-2/CD86; B7-H1/PD-L1; B7-H2; B7-H3; B7-H4; B7-H6; B7-H7; BTLA/CD272; CD28; CTLA-4; Gi24/VISTA/B7-H5; ICOS/CD278; PD-1; PD-L2/B7-DC; PDCD6); 4-1BB/TNFSF9/CD137; 4-1BB Ligand/TNFSF9; BAFF/BLyS/TNFSF13B; BAFF R/TNFRSF13C; CD27/TNFRSF7; CD27 Ligand/TNFSF7; CD30/TNFRSF8; CD30 Ligand/TNFSF8; CD40/TNFRSF5; CD40/TNFSF5; CD40 Ligand/TNFSF5; DR3/TNFRSF25; GITR/TNFRSF18; GITR Ligand/TNFSF18; HVEM/TNFRSF14; LIGHT/TNFSF14; Lymphotoxin-alpha/TNF-beta; OX40/TNFRSF4; OX40 Ligand/TNFSF4; RELT/TNFRSF19L; TACI/TNFRSF13B; TL1A/TNFSF15; TNF-alpha; TNF RII/TNFRSF1B); 2B4/CD244/SLAMF4; BLAME/SLAMF8; CD2; CD2F-10/SLAMF9; CD48/SLAMF2; CD58/LFA-3; CD84/SLAMF5; CD229/SLAMF3; CRACC/SLAMF7; NTB-

A/SLAMF6; SLAM/CD150); CD2; CD7; CD53; CD82/Kai-1; CD90/Thy1; CD96; CD160; CD200; CD300a/LMIR1; HLA Class I; HLA-DR; Ikaros; Integrin alpha 4/CD49d; Integrin alpha 4 beta 1; Integrin alpha 4 beta 7/LPAM-1; LAG-3; TCL1A; TCL1B; CRTAM; DAP12; Dectin-1/CLEC7A; DPPIV/CD26; EphB6; TIM-1/KIM-1/HAVCR; TIM-4; TSLP; TSLP R; lymphocyte function associated antigen-1 (LFA-1); NKG2C, a CD3 zeta domain, an immunoreceptor tyrosine-based activation motif (ITAM), CD27, CD28, 4-1BB, CD134/OX40, CD30, CD40, PD-1, ICOS, lymphocyte function-associated antigen-1 (LFA-1), CD2, CD7, LIGHT, NKG2C, B7-H3, a ligand that specifically binds with CD83, or functional fragment thereof.

**[0360]** In some embodiments, the CAR comprises a CD3 zeta domain or an immunoreceptor tyrosine-based activation motif (ITAM), or functional variant thereof. In some embodiments, the CAR comprises (i) a CD3 zeta domain, or an immunoreceptor tyrosine-based activation motif (ITAM), or functional variant thereof; and (ii) a CD28 domain, or a 4-1BB domain, or functional variant thereof. In some embodiments, the CAR comprises a (i) a CD3 zeta domain, or an immunoreceptor tyrosine-based activation motif (ITAM), or functional variant thereof; (ii) a CD28 domain or functional variant thereof; and (iii) a 4-1BB domain, or a CD134 domain, or functional variant thereof. In some embodiments, the CAR comprises (i) a CD3 zeta domain, or an immunoreceptor tyrosine-based activation motif (ITAM), or functional variant thereof; (ii) a CD28 domain, or a 4-1BB domain, or functional variant thereof, and/or (iii) a 4-1BB domain, or a CD134 domain, or functional variant thereof. In some embodiments, the CAR comprises a (i) a CD3 zeta domain, or an immunoreceptor tyrosine-based activation motif (ITAM), or functional variant thereof; (ii) a CD28 domain or functional variant thereof; (iii) a 4-1BB domain, or a CD134 domain, or functional variant thereof; and (iv) a cytokine or costimulatory ligand transgene.

**[0361]** In some embodiments, the CAR further comprises one or more spacers, e.g., wherein the spacer is a first spacer between the antigen binding domain and the transmembrane domain. In some embodiments, the first spacer includes at least a portion of an immunoglobulin constant region or variant or modified version thereof. In some embodiments, the spacer is a second spacer between the transmembrane domain and a signaling domain. In some embodiments, the second spacer is an oligopeptide, e.g., wherein the oligopeptide comprises glycine-serine doublets.

**[0362]** In some embodiments the exogenous agent is or comprises a CAR, e.g., a first generation CAR or a nucleic acid encoding a first generation CAR. In some embodiments, a first generation CAR comprises an antigen binding domain, a transmembrane domain, and signaling domain. In some embodiments a signaling domain mediates downstream signaling during T cell activation.

**[0363]** In some embodiments the exogenous agent is or comprises a second generation CAR or a nucleic acid encoding a second generation CAR. In some embodiments a second generation CAR comprises an antigen binding domain, a transmembrane domain, and two signaling domains. In some embodiments a signaling domain mediates downstream signaling during T cell activation. In some

embodiments a signaling domain is a costimulatory domain. In some embodiments, a costimulatory domain enhances cytokine production, CAR T cell proliferation, and or CAR T cell persistence during T cell activation.

**[0364]** In some embodiments the exogenous agent is or comprises a third generation CAR or a nucleic acid encoding a third generation CAR. In some embodiments, a third generation CAR comprises an antigen binding domain, a transmembrane domain, and at least three signaling domains. In some embodiments a signaling domain mediates downstream signaling during T cell activation. In some embodiments a signaling domain is a costimulatory domain. In some embodiments, a costimulatory domain enhances cytokine production, CAR T cell proliferation, and or CAR T cell persistence during T cell activation. In some embodiments, a third generation CAR comprises at least two costimulatory domains. In some embodiments, the at least two costimulatory domains are not the same.

**[0365]** In some embodiments the exogenous is or comprises a fourth generation CAR or a nucleic acid encoding a fourth generation CAR. In some embodiments a fourth generation CAR comprises an antigen binding domain, a transmembrane domain, and at least two, three, or four signaling domains. In some embodiments a signaling domain mediates downstream signaling during T cell activation. In some embodiments a signaling domain is a costimulatory domain. In some embodiments, a costimulatory domain enhances cytokine production, CAR T cell proliferation, and or CAR T cell persistence during T cell activation.

**[0366]** In some embodiments, a first, second, third, or fourth generation CAR further comprises a domain which upon successful signaling of the CAR induces expression of a cytokine gene. In some embodiments, a cytokine gene is endogenous or exogenous to a target cell comprising a CAR which comprises a domain which upon successful signaling of the CAR induces expression of a cytokine gene. In some embodiments a cytokine gene encodes a pro-inflammatory cytokine. In some embodiments a cytokine gene encodes IL-1, IL-2, IL-9, IL-12, IL-18, TNF, or IFN-gamma, or functional fragment thereof. In some embodiments a domain which upon successful signaling of the CAR induces expression of a cytokine gene is or comprises a transcription factor or functional domain or fragment thereof. In some embodiments a domain which upon successful signaling of the CAR induces expression of a cytokine gene is or comprises a transcription factor or functional domain or fragment thereof. In some embodiments a transcription factor or functional domain or fragment thereof is or comprises a nuclear factor of activated T cells (NFAT), an NF-kB, or functional domain or fragment thereof. See, e.g., Zhang. C. et al., Engineering CAR-T cells. Biomarker Research. 5:22 (2017); WO 2016126608; Sha, H. et al. Chimaeric antigen receptor T-cell therapy for tumour immunotherapy. Bioscience Reports Jan 27, 2017, 37 (1).

**[0367]** In some embodiments, a CAR antigen binding domain is or comprises an antibody or antigen-binding portion thereof. In some embodiments, a CAR antigen binding domain is or comprises an scFv or Fab. In some embodiments a CAR antigen binding domain comprises an scFv or Fab

fragment of a T-cell alpha chain antibody; T-cell β chain antibody; T-cell γ chain antibody; T-cell δ chain antibody; CCR7 antibody; CD3 antibody; CD4 antibody; CD5 antibody; CD7 antibody; CD8 antibody; CD11b antibody; CD11c antibody; CD16 antibody; CD19 antibody; CD20 antibody; CD21 antibody; CD22 antibody; CD25 antibody; CD28 antibody; CD34 antibody; CD35 antibody; CD40 antibody; CD45RA antibody; CD45RO antibody; CD52 antibody; CD56 antibody; CD62L antibody; CD68 antibody; CD80 antibody; CD95 antibody; CD117 antibody; CD127 antibody; CD133 antibody; CD137 (4-1 BB) antibody; CD163 antibody; F4/80 antibody; IL-4Ra antibody; Sca-1 antibody; CTLA-4 antibody; GITR antibody GARP antibody; LAP antibody; granzyme B antibody; LFA-1 antibody; MR1 antibody; uPAR antibody; or transferrin receptor antibody.

**[0368]** In some embodiments, an antigen binding domain binds to a cell surface antigen of a cell. In some embodiments, a cell surface antigen is characteristic of one type of cell. In some embodiments, a cell surface antigen is characteristic of more than one type of cell.

**[0369]** In some embodiments a CAR antigen binding domain binds a cell surface antigen characteristic of a T cell. In some embodiments, an antigen characteristic of a T cell may be a cell surface receptor, a membrane transport protein (e.g., an active or passive transport protein such as, for example, an ion channel protein, a pore-forming protein, etc.), a transmembrane receptor, a membrane enzyme, and/or a cell adhesion protein characteristic of a T cell. In some embodiments, an antigen characteristic of a T cell may be a G protein-coupled receptor, receptor tyrosine kinase, tyrosine kinase associated receptor, receptor-like tyrosine phosphatase, receptor serine/ threonine kinase, receptor guanylyl cyclase, or histidine kinase associated receptor.

**[0370]** In some embodiments, an antigen characteristic of a T cell may be a T cell receptor. In some embodiments, a T cell receptor may be AKT1; AKT2; AKT3; ATF2; BCL10; CALM1; CD3D (CD3δ); CD3E (CD3ε); CD3G (CD3γ); CD4; CD8; CD28; CD45; CD80 (B7-1); CD86 (B7-2); CD247 (CD3ζ); CTLA4 (CD152); ELK1; ERK1 (MAPK3); ERK2; FOS; FYN; GRAP2 (GADS); GRB2; HLA-DRA; HLA-DRB1; HLA-DRB3; HLA-DRB4; HLA-DRB5; HRAS; IKBKA (CHUK); IKBKB; IKBKE; IKBKG (NEMO); IL2; ITPR1; ITK; JUN; KRAS2; LAT; LCK; MAP2K1 (MEK1); MAP2K2 (MEK2); MAP2K3 (MKK3); MAP2K4 (MKK4); MAP2K6 (MKK6); MAP2K7 (MKK7); MAP3K1 (MEKK1); MAP3K3; MAP3K4; MAP3K5; MAP3K8; MAP3K14 (NIK); MAPK8 (JNK1); MAPK9 (JNK2); MAPK10 (JNK3); MAPK11 (p38β); MAPK12 (p38γ); MAPK13 (p38δ); MAPK14 (p38α); NCK; NFAT1; NFAT2; NFKB1; NFKB2; NFKBIA; NRAS; PAK1; PAK2; PAK3; PAK4; PIK3C2B; PIK3C3 (VPS34); PIK3CA; PIK3CB; PIK3CD; PIK3R1; PKCA; PKCB; PKCM; PKCQ; PLCY1; PRF1 (Perforin); PTEN; RAC1; RAF1; RELA; SDF1; SHP2; SLP76; SOS; SRC; TBK1; TCRA; TEC; TRAF6; VAV1; VAV2; or ZAP70.

**[0371]** In some embodiments a CAR comprises a signaling domain which is a costimulatory domain. In some embodiments a CAR comprises a second costimulatory domain. In some embodiments a CAR comprises at least two costimulatory domains. In some embodiments a CAR comprises at least

three costimulatory domains. In some embodiments a CAR comprises a costimulatory domain selected from one or more of CD27, CD28, 4-1BB, CD134/OX40, CD30, CD40, PD-1, ICOS, lymphocyte function-associated antigen-1 (LFA-1), CD2, CD7, LIGHT, NKG2C, B7-H3, a ligand that specifically binds with CD83.

**[0372]** In some embodiments, the CAR comprises a CD3 zeta domain or an immunoreceptor tyrosine-based activation motif (ITAM), or functional variant thereof. In some embodiments, the CAR comprises (i) a CD3 zeta domain, or an immunoreceptor tyrosine-based activation motif (ITAM), or functional variant thereof; and (ii) a CD28 domain, or a 4-1BB domain, or functional variant thereof. In some embodiments, the CAR comprises a (i) a CD3 zeta domain, or an immunoreceptor tyrosine-based activation motif (ITAM), or functional variant thereof; (ii) a CD28 domain or functional variant thereof; and (iii) a 4-1BB domain, or a CD134 domain, or functional variant thereof. In some embodiments, the CAR comprises (i) a CD3 zeta domain, or an immunoreceptor tyrosine-based activation motif (ITAM), or functional variant thereof; (ii) a CD28 domain, or a 4-1BB domain, or functional variant thereof, and/or (iii) a 4-1BB domain, or a CD134 domain, or functional variant thereof. In some embodiments, the CAR comprises a (i) a CD3 zeta domain, or an immunoreceptor tyrosine-based activation motif (ITAM), or functional variant thereof; (ii) a CD28 domain or functional variant thereof; (iii) a 4-1BB domain, or a CD134 domain, or functional variant thereof; and (iv) a cytokine or costimulatory ligand transgene.

**[0373]** In certain embodiments, the intracellular signaling domain comprises a CD28 transmembrane and signaling domain linked to a CD3 (e.g., CD3-zeta) intracellular domain. In some embodiments, the intracellular signaling domain comprises a chimeric CD28 and CD137 (4-1BB, TNFRSF9) co-stimulatory domains, linked to a CD3 zeta intracellular domain.

**[0374]** In some embodiments, the CAR encompasses one or more, e.g., two or more, costimulatory domains and an activation domain, e.g., primary activation domain, in the cytoplasmic portion. Exemplary CARs include intracellular components of CD3-zeta, CD28, and 4-1BB.

**[0375]** In some embodiments the intracellular signaling domain includes intracellular components of a 4-1BB signaling domain and a CD3-zeta signaling domain. In some embodiments, the intracellular signaling domain includes intracellular components of a CD28 signaling domain and a CD3zeta signaling domain.

**[0376]** In some embodiments, the CAR comprises an extracellular antigen binding domain (e.g., antibody or antibody fragment, such as an scFv) that binds to an antigen (e.g. tumor antigen), a spacer (e.g. containing a hinge domain, such as any as described herein), a transmembrane domain (e.g. any as described herein), and an intracellular signaling domain (e.g. any intracellular signaling domain, such as a primary signaling domain or costimulatory signaling domain as described herein). In some embodiments, the intracellular signaling domain is or includes a primary cytoplasmic signaling domain.

In some embodiments, the intracellular signaling domain additionally includes an intracellular signaling domain of a costimulatory molecule (e.g., a costimulatory domain).

**[0377]** In some embodiments, the CAR contains one or more domains that combine an antigen- or ligand-binding domain (e.g. antibody or antibody fragment) that provides specificity for a desired antigen (e.g., tumor antigen) with intracellular signaling domains. In some embodiments, the intracellular signaling domain is a stimulating or an activating intracellular domain portion, such as a T cell stimulating or activating domain, providing a primary activation signal or a primary signal. In some embodiments, the intracellular signaling domain contains or additionally contains a costimulatory signaling domain to facilitate effector functions. In some embodiments, chimeric receptors when genetically engineered into immune cells can modulate T cell activity, and, in some cases, can modulate T cell differentiation or homeostasis, thereby resulting in genetically engineered cells with improved longevity, survival and/or persistence in vivo, such as for use in adoptive cell therapy methods.

**[0378]** Exemplary antigen receptors, including CARs, and methods for engineering and introducing such receptors into cells, include those described, for example, in W0200014257, WO2013126726, WO2012/129514, WO2014031687, WO2013/166321, WO2013/071154, W02013/123061, U.S. patent app. Pub. Nos. US2002131960, US2013287748, US20130149337, U.S. Patent Nos. 6,451,995, 7,446,190, 8,252,592, 8,339,645, 8,398,282, 7,446,179, 6,410,319, 7,070,995, 7,265,209, 7,354,762, 7,446,191, 8,324,353, and 8,479,118, and European patent app. No. EP2537416, and/or those described by Sadelain et al., Cancer Discov. 2013 April; 3(4): 388-398; Davila et al. (2013) PLoS ONE 8(4): e61338; Turtle et al., Curr. Opin. Immunol., 2012 October; 24(5): 633-39; Wu et al., Cancer, 2012 March 18(2): 160-75. In some aspects, the antigen receptors include a CAR as described in U.S. Patent No.: 7,446,190, and those described in WO/2014055668. Examples of the CARs include CARs as disclosed in any of the aforementioned publications, such as WO2014031687, US 8,339,645, US 7,446,179, US 2013/0149337, US 7,446,190, US 8,389,282, Kochenderfer et al., (2013) Nature Reviews Clinical Oncology, 10, 267-276; Wang et al. (2012) J. Immunother. 35(9): 689-701; and Brentjens et al., Sci Transl Med. 2013 5(177). See also WO2014031687, US 8,339,645, US 7,446,179, US 2013/0149337, US 7,446,190, and US 8,389,282.The recombinant receptors, such as CARs, generally include an extracellular antigen binding domain, such as a portion of an antibody molecule, generally a variable heavy (VH) chain region and/or variable light (VL) chain region of the antibody, e.g., an scFv antibody fragment. In some embodiments, the antigen binding domain of the CAR molecule comprises an antibody, an antibody fragment, an scFv, a Fv, a Fab, a $(Fab')_2$, a single domain antibody (SdAb), a VH or VL domain, or a camelid VHH domain.

**[0379]** In some embodiments, a CAR antigen binding domain is or comprises an antibody or antigen-binding portion thereof. In some embodiments, a CAR antigen binding domain is or comprises an scFv or Fab. In some embodiments, a CAR antigen binding domain comprises an scFv or Fab fragment of a CD19 antibody; CD22 antibody; T-cell alpha chain antibody; T-cell β chain antibody; T-

cell γ chain antibody; T-cell δ chain antibody; CCR7 antibody; CD3 antibody; CD4 antibody; CD5 antibody; CD7 antibody; CD8 antibody; CD11b antibody; CD11c antibody; CD16 antibody; CD20 antibody; CD21 antibody; CD25 antibody; CD28 antibody; CD34 antibody; CD35 antibody; CD40 antibody; CD45RA antibody; CD45RO antibody; CD52 antibody; CD56 antibody; CD62L antibody; CD68 antibody; CD80 antibody; CD95 antibody; CD117 antibody; CD127 antibody; CD133 antibody; CD137 (4-1 BB) antibody; CD163 antibody; F4/80 antibody; IL-4Ra antibody; Sca-1 antibody; CTLA-4 antibody; GITR antibody GARP antibody; LAP antibody; granzyme B antibody; LFA-1 antibody; MR1 antibody; uPAR antibody; or transferrin receptor antibody.

**[0380]** In some embodiments, a CAR comprises a signaling domain which is a costimulatory domain. In some embodiments, a CAR comprises a second costimulatory domain. In some embodiments, a CAR comprises at least two costimulatory domains. In some embodiments, a CAR comprises at least three costimulatory domains. In some embodiments, a CAR comprises a costimulatory domain selected from one or more of CD27, CD28, 4-1BB, CD134/OX40, CD30, CD40, PD-1, ICOS, lymphocyte function-associated antigen-1 (LFA-1), CD2, CD7, LIGHT, NKG2C, B7-H3, a ligand that specifically binds with CD83. In some embodiments, if a CAR comprises two or more costimulatory domains, two costimulatory domains are different. In some embodiments, if a CAR comprises two or more costimulatory domains, two costimulatory domains are the same.

**[0381]** Examples of exemplary components of a CAR are described in **Table 3C**. In provided aspects, the sequences of each component in a CAR can include any combination listed in Table 3C.

Table 3C: CAR components and Exemplary Sequences

| **Component** | **Sequence** | **SEQ ID NO** |
|---|---|---|
| *Extracellular binding domain* | | |
| Anti-CD19 scFv (FMC63) | DIQMTQTTSSLSASLGDRVTISCRASQDISKYLN WYQQKPDGTVKLLIYHTSRLHSGVPSRFSGSGSG TDYSLTISNLEQEDIATYFCQQGNTLPYTFGGGT KLEITGSTSGSGKPGSGEGSTKGEVKLQESGPGL VAPSQSLSVTCTVSGVSLPDYGVSWIRQPPRKGL EWLGVIWGSETTYYNSALKSRLTIIKDNSKSQVF LKMNSLQTDDTAIYYCAKHYYYGGSYAMDYW GQGTSVTVSS | 274 |
| Anti-CD19 scFv (FMC63) | DIQMTQTTSSLSASLGDRVTISCRASQDISKYLN WYQQKPDGTVKLLIYHTSRLHSGVPSRFSGSGSG TDYSLTISNLEQEDIATYFCQQGNTLPYTFGGGT KLEITGGGGSGGGGSGGGGSEVKLQESGPGLVA PSQSLSVTCTVSGVSLPDYGVSWIRQPPRKGLEW LGVIWGSETTYYNSALKSRLTIIKDNSKSQVFLK MNSLQTDDTAIYYCAKHYYYGGSYAMDYWGQ GTSVTVSS | 275 |
| *Spacer (e.g. hinge)* | | |
| IgG4 Hinge | ESKYGPPCPPCP | 276 |

| | | |
|---|---|---|
| CD8 Hinge | TTTPAPRPPTPAPTIASQPLSLRPE | 277 |
| CD28 | IEVMYPPPYLDNEKSNGTIIHVKGKHLCPSPLFPG PSKP | 278 |
| *Transmembrane* | | |
| CD8 | ACRPAAGGAVHTRGLDFACDIYIWAPLAGTCGV LLLSLVITLYC | 279 |
| CD28 | FWVLVVVGGVLACYSLLVTVAFIIFWV | 280 |
| CD28 | FWVLVVVGGVLACYSLLVTVAFIIFWV | 281 |
| *Costimulatory domain* | | |
| CD28 | RSKRSRLLHSDYMNMTPRRPGPTRKHYQPYAPP RDFAAYRS | 282 |
| 4-1BB | KRGRKKLLYIFKQPFMRPVQTTQEEDGCSCRFPE EEEGGCEL | 283 |
| *Primary Signaling Domain* | | |
| CD3zeta | RVKFSRSADAPAYQQGQNQLYNELNLGRREEYD VLDKRRGRDPEMGGKPRRKNPQEGLYNELQKD KMAEAYSEIGMKGERRRGKGHDGLYQGLSTAT KDTYDALHMQALPPR | 284 |
| CD3zeta | RVKFSRSADAPAYKQGQNQLYNELNLGRREEYD VLDKRRGRDPEMGGKPRRKNPQEGLYNELQKD KMAEAYSEIGMKGERRRGKGHDGLYQGLSTAT KDTYDALHMQALPPR | 285 |

**[0382]** In some embodiments, the CAR further comprises one or more spacers, e.g., wherein the spacer is a first spacer between the antigen binding domain and the transmembrane domain. In some embodiments, the first spacer includes at least a portion of an immunoglobulin constant region or variant or modified version thereof. In some embodiments, the spacer is a second spacer between the transmembrane domain and a signaling domain. In some embodiments, the second spacer is an oligopeptide, e.g., wherein the oligopeptide comprises glycine-serine doublets.

**[0383]** In addition to the CARs described herein, various chimeric antigen receptors and nucleotide sequences encoding the same are known in the art and would be suitable for fusosomal delivery and reprogramming of target cells in vivo and in vitro as described herein. See, e.g., WO2013040557; WO2012079000; WO2016030414; Smith T, et al., Nature Nanotechnology. 2017. DOI: 10.1038/NNANO.2017.57, the disclosures of which are herein incorporated by reference.

**[0384]** In some embodiments, the antigen targeted by the receptor is a polypeptide. In some embodiments, it is a carbohydrate or other molecule. In some embodiments, the antigen is selectively expressed or overexpressed on cells of the disease or condition, e.g., the tumor or pathogenic cells, as compared to normal or non-targeted cells or tissues. In other embodiments, the antigen is expressed on normal cells and/or is expressed on the engineered cells.

**[0385]** In some embodiments, the antigen targeted by the receptor includes antigens associated with a B cell malignancy, such as any of a number of known B cell markers. In some embodiments, the

antigen targeted by the receptor is CD20, CD19, CD22, ROR1, CD45, CD47, CD21, CD5, CD33, Igkappa, Iglambda, CD79a, CD79b or CD30.

**[0386]** In some embodiments, the CAR binds to CD19. In some embodiments, the CAR binds to CD22. In some embodiments, the CAR binds to CD19 and CD22. In some embodiments, the CAR is selected from the group consisting of a first generation CAR, a second generation CAR, a third generation CAR, and a fourth generation CAR. In some embodiments, the CAR includes a single binding domain that binds to a single target antigen. In some embodiments, the CAR includes a single binding domain that binds to more than one target antigen, *e.g.*, 2, 3, or more target antigens. In some embodiments, the CAR includes two binding domains such that each binding domain binds to a different target antigens. In some embodiments, the CAR includes two binding domains such that each binding domain binds to the same target antigen. Detailed descriptions of exemplary CARs including CD19-specific, CD22-specific and CD19/CD22-bispecific CARs can be found in WO2012/079000, WO2016/149578 and WO2020/014482, the disclosures including the sequence listings and figures are incorporated herein by reference in their entirety.

**[0387]** In some embodiments, the chimeric antigen receptor includes an extracellular portion containing an antibody or antibody fragment. In some aspects, the chimeric antigen receptor includes an extracellular portion containing the antibody or fragment and an intracellular signaling domain. In some embodiments, the antibody or fragment includes an scFv.

**[0388]** In some embodiments, the antigen targeted by the antigen-binding domain is CD19. In some aspects, the antigen-binding domain of the recombinant receptor, e.g., CAR, and the antigen-binding domain binds, such as specifically binds or specifically recognizes, a CD19, such as a human CD19. In some embodiments, the scFv contains a VH and a VL derived from an antibody or an antibody fragment specific to CD19. In some embodiments, the antibody or antibody fragment that binds CD19 is a mouse derived antibody such as FMC63 and SJ25C1. In some embodiments, the antibody or antibody fragment is a human antibody, e.g., as described in U.S. Patent Publication No. US 2016/0152723.

**[0389]** In some embodiments, the antigen is CD19. In some embodiments, the scFv contains a VH and a VL derived from an antibody or an antibody fragment specific to CD 19. In some embodiments, the antibody or antibody fragment that binds CD 19 is a mouse derived antibody such as FMC63 and SJ25C1. In some embodiments, the antibody or antibody fragment is a human antibody, e.g., as described in U.S. Patent Publication No. US 2016/0152723.

**[0390]** In some embodiments, the scFv is derived from FMC63. FMC63 generally refers to a mouse monoclonal IgGl antibody raised against Naim-1 and -16 cells expressing CD19 of human origin (Fing, N. R., et al. (1987). Leucocyte typing III. 302).

**[0391]** In some embodiments, the antibody portion of the recombinant receptor, e.g., CAR, further includes spacer between the transmembrane domain and extracellular antigen binding domain. In some embodiments, the spacer includes at least a portion of an immunoglobulin constant region, such as a

hinge region, e.g., an IgG4 hinge region, and/or a CH1/CL and/or Fc region. In some embodiments, the constant region or portion is of a human IgG, such as IgG4 or IgGl. In some aspects, the portion of the constant region serves as a spacer region between the antigen-recognition component, e.g., scFv, and transmembrane domain. The spacer can be of a length that provides for increased responsiveness of the cell following antigen binding, as compared to in the absence of the spacer. Exemplary spacers include, but are not limited to, those described in Hudecek et al. (2013) Clin. Cancer Res., 19:3153, WO2014031687, U.S. Patent No. 8,822,647 or published app. No. US 2014/0271635. In some embodiments, the constant region or portion is of a human IgG, such as IgG4 or IgGl.

**[0392]** In some embodiments, the antigen receptor comprises an intracellular domain linked directly or indirectly to the extracellular domain. In some embodiments, the chimeric antigen receptor includes a transmembrane domain linking the extracellular domain and the intracellular signaling domain. In some embodiments, the intracellular signaling domain comprises an IT AM. For example, in some aspects, the antigen recognition domain (e.g. extracellular domain) generally is linked to one or more intracellular signaling components, such as signaling components that mimic activation through an antigen receptor complex, such as a TCR complex, in the case of a CAR, and/or signal via another cell surface receptor. In some embodiments, the chimeric receptor comprises a transmembrane domain linked or fused between the extracellular domain (e.g. scFv) and intracellular signaling domain. Thus, in some embodiments, the antigen-binding component (e.g., antibody) is linked to one or more transmembrane and intracellular signaling domains.

**[0393]** In one embodiment, a transmembrane domain that naturally is associated with one of the domains in the receptor, e.g., CAR, is used. In some instances, the transmembrane domain is selected or modified by amino acid substitution to avoid binding of such domains to the transmembrane domains of the same or different surface membrane proteins to minimize interactions with other members of the receptor complex.

**[0394]** In some embodiments, the CAR transmembrane domain comprises at least a transmembrane region of the alpha, beta or zeta chain of a T cell receptor, CD28, CD3 epsilon, CD45, CD4, CD5, CD8, CD9, CD16, CD22, CD33, CD37, CD64, CD80, CD86, CD134, CD137, CD154, or functional variant thereof. In some embodiments, the transmembrane domain comprises at least a transmembrane region(s) of CD8α, CD8β, 4-1BB/CD137, CD28, CD34, CD4, FcεRIγ, CD16, OX40/CD134, CD3ζ, CD3ε, CD3γ, CD3δ, TCRα, TCRβ, TCRζ, CD32, CD64, CD64, CD45, CD5, CD9, CD22, CD37, CD80, CD86, CD40, CD40L/CD154, VEGFR2, FAS, and FGFR2B, or functional variant thereof. The transmembrane domain in some embodiments is derived either from a natural or from a synthetic source. Where the source is natural, the domain in some aspects is derived from any membrane-bound or transmembrane protein. Transmembrane regions include those derived from (i.e. comprise at least the transmembrane region(s) of) the alpha, beta or zeta chain of the T-cell receptor, CD28, CD3 epsilon, CD45, CD4, CD5, CD8, CD9, CD16, CD22, CD33, CD37, CD64, CD80, CD86, CD 134, CD 137, CD 154. Alternatively the

transmembrane domain in some embodiments is synthetic. In some aspects, the synthetic transmembrane domain comprises predominantly hydrophobic residues such as leucine and valine. In some aspects, a triplet of phenylalanine, tryptophan and valine will be found at each end of a synthetic transmembrane domain. In some embodiments, the linkage is by linkers, spacers, and/or transmembrane domain(s). In some aspects, the transmembrane domain contains a transmembrane portion of CD28.

**[0395]** In some embodiments, the extracellular domain and transmembrane domain can be linked directly or indirectly. In some embodiments, the extracellular domain and transmembrane are linked by a spacer, such as any described herein. In some embodiments, the receptor contains extracellular portion of the molecule from which the transmembrane domain is derived, such as a CD28 extracellular portion.

**[0396]** Among the intracellular signaling domains are those that mimic or approximate a signal through a natural antigen receptor, a signal through such a receptor in combination with a costimulatory receptor, and/or a signal through a costimulatory receptor alone. In some embodiments, a short oligo- or polypeptide linker, for example, a linker of between 2 and 10 amino acids in length, such as one containing glycines and serines, e.g., glycine-serine doublet, is present and forms a linkage between the transmembrane domain and the cytoplasmic signaling domain of the CAR.

**[0397]** T cell activation is in some aspects described as being mediated by two classes of cytoplasmic signaling sequences: those that initiate antigen-dependent primary activation through the TCR (primary cytoplasmic signaling sequences), and those that act in an antigen-independent manner to provide a secondary or co-stimulatory signal (secondary cytoplasmic signaling sequences). In some aspects, the CAR includes one or both of such signaling components.

**[0398]** The receptor, e.g., the CAR, generally includes at least one intracellular signaling component or components. In some aspects, the CAR includes a primary cytoplasmic signaling sequence that regulates primary activation of the TCR complex. Primary cytoplasmic signaling sequences that act in a stimulatory manner may contain signaling motifs which are known as immunoreceptor tyrosine -based activation motifs or IT AMs. Examples of IT AM containing primary cytoplasmic signaling sequences include those derived from CD3 zeta chain, FcR gamma, CD3 gamma, CD3 delta and CD3 epsilon. In some embodiments, cytoplasmic signaling molecule(s) in the CAR contain(s) a cytoplasmic signaling domain, portion thereof, or sequence derived from CD3 zeta.

**[0399]** In some embodiments, the receptor includes an intracellular component of a TCR complex, such as a TCR CD3 chain that mediates T-cell activation and cytotoxicity, e.g., CD3 zeta chain. Thus, in some aspects, the antigen-binding portion is linked to one or more cell signaling modules. In some embodiments, cell signaling modules include CD3 transmembrane domain, CD3 intracellular signaling domains, and/or other CD transmembrane domains. In some embodiments, the intracellular component is or includes a CD3-zeta intracellular signaling domain. In some embodiments, the intracellular component is or includes a signaling domain from Fc receptor gamma chain. In some embodiments, the receptor, e.g., CAR, includes the intracellular signaling domain and further includes a portion, such as a

transmembrane domain and/or hinge portion, of one or more additional molecules such as CD8, CD4, CD25, or CD 16. For example, in some aspects, the CAR or other chimeric receptor is a chimeric molecule of CD3-zeta (CD3-z) or Fc receptor and a portion of one of CD8, CD4, CD25 or CD16.

**[0400]** In some embodiments, upon ligation of the CAR or other chimeric receptor, the cytoplasmic domain or intracellular signaling domain of the receptor activates at least one of the normal effector functions or responses of the immune cell, e.g., T cell engineered to express the CAR. For example, in some contexts, the CAR induces a function of a T cell such as cytolytic activity or T-helper activity, such as secretion of cytokines or other factors. In some embodiments, a truncated portion of an intracellular signaling domain of an antigen receptor component or costimulatory molecule is used in place of an intact immunostimulatory chain, for example, if it transduces the effector function signal. In some embodiments, the intracellular signaling domain or domains include the cytoplasmic sequences of the T cell receptor (TCR), and in some aspects also those of co-receptors that in the natural context act in concert with such receptors to initiate signal transduction following antigen receptor engagement.

**[0401]** In the context of a natural TCR, full activation generally requires not only signaling through the TCR, but also a costimulatory signal. Thus, in some embodiments, to promote full activation, a component for generating secondary or co-stimulatory signal is also included in the CAR. In other embodiments, the CAR does not include a component for generating a costimulatory signal. In some aspects, an additional CAR is expressed in the same cell and provides the component for generating the secondary or costimulatory signal.

**[0402]** In some embodiments, the chimeric antigen receptor contains an intracellular domain of a T cell costimulatory molecule. In some embodiments, the CAR includes a signaling domain and/or transmembrane portion of a costimulatory receptor, such as CD28, 4-1BB, OX40, DAP10, and ICOS. In some aspects, the same CAR includes both the activating and costimulatory components. In some embodiments, the chimeric antigen receptor contains an intracellular domain derived from a T cell costimulatory molecule or a functional variant thereof, such as between the transmembrane domain and intracellular signaling domain. In some aspects, the T cell costimulatory molecule is CD28 or 41BB. In some aspects, the T cell costimulatory molecule is 41BB.

**[0403]** In some embodiments, the activating domain is included within one CAR, whereas the costimulatory component is provided by another CAR recognizing another antigen. In some embodiments, the CARs include activating or stimulatory CARs, costimulatory CARs, both expressed on the same cell (see WO2014/055668). In some aspects, the cells include one or more stimulatory or activating CAR and/or a costimulatory CAR. In some embodiments, the cells further include inhibitory CARs (iCARs, see Fedorov et al., Sci. Transl. Medicine, 5(215) (December, 2013), such as a CAR recognizing an antigen other than the one associated with and/or specific for the disease or condition whereby an activating signal delivered through the disease-targeting CAR is diminished or inhibited by binding of the inhibitory CAR to its ligand, e.g., to reduce off-target effects.

**[0404]** In certain embodiments, the intracellular signaling domain comprises a CD28 transmembrane and signaling domain linked to a CD3 (e.g., CD3-zeta) intracellular domain. In some embodiments, the intracellular signaling domain comprises a chimeric CD28 and CD137 (4-1BB, TNFRSF9) co-stimulatory domains, linked to a CD3 zeta intracellular domain.

**[0405]** In some embodiments, the CAR encompasses one or more, e.g., two or more, costimulatory domains and an activation domain, e.g., primary activation domain, in the cytoplasmic portion. Exemplary CARs include intracellular components of CD3-zeta, CD28, and 4-1BB.

**[0406]** In some embodiments the intracellular signaling domain includes intracellular components of a 4-1BB signaling domain and a CD3-zeta signaling domain. In some embodiments, the intracellular signaling domain includes intracellular components of a CD28 signaling domain and a CD3zeta signaling domain.

**[0407]** In some embodiments, a CD19 specific CAR includes an anti-CD19 single-chain antibody fragment (scFv), a transmembrane domain such as one derived from human CD8α, a 4-1BB (CD137) co-stimulatory signaling domain, and a CD3ζ signaling domain. In some embodiments, a CD22 specific CAR includes an anti-CD22 scFv, a transmembrane domain such as one derived from human CD8α, a 4-1BB (CD137) co-stimulatory signaling domain, and a CD3ζ signaling domain. In some embodiments, a CD19/CD22-bispecific CAR includes an anti-CD19 scFv, an anti-CD22 scFv, a transmembrane domain such as one derived from human CD8α, a 4-1BB (CD137) co-stimulatory signaling domain, and a CD3ζ signaling domain.

**[0408]** In some embodiments, the CAR comprises a commercial CAR construct carried by a T cell. Non-limiting examples of commercial CAR-T cell based therapies include brexucabtagene autoleucel (TECARTUS®), axicabtagene ciloleucel (YESCARTA®), idecabtagene vicleucel (ABECMA®), lisocabtagene maraleucel (BREYANZI®), tisagenlecleucel (KYMRIAH®), Descartes-08 and Descartes-11 from Cartesian Therapeutics, CTL110 from Novartis, P-BMCA-101 from Poseida Therapeutics, AUTO4 from Autolus Limited, UCARTCS from Cellectis, PBCAR19B and PBCAR269A from Precision Biosciences, FT819 from Fate Therapeutics, and CYAD-211 from Clyad Oncology.

**[0409]** In some embodiments, the antigen binding domain targets an antigen characteristic of an autoimmune or inflammatory disorder. In some embodiments, the ABD binds an antigen associated with an autoimmune or inflammatory disorder. In some instances, the antigen is expressed by a cell associated with an autoimmune or inflammatory disorder. In some embodiments, the autoimmune or inflammatory disorder is selected from chronic graft-vs-host disease (GVHD), lupus, arthritis, immune complex glomerulonephritis, goodpasture syndrome, uveitis, hepatitis, systemic sclerosis or scleroderma, type I diabetes, multiple sclerosis, cold agglutinin disease, Pemphigus vulgaris, Grave's disease, autoimmune hemolytic anemia, Hemophilia A, Primary Sjogren's Syndrome, thrombotic thrombocytopenia purrpura, neuromyelits optica, Evan's syndrome, IgM mediated neuropathy, cryoglobulinemia, dermatomyositis, idiopathic thrombocytopenia, ankylosing spondylitis, bullous

pemphigoid, acquired angioedema, chronic urticarial, antiphospholipid demyelinating polyneuropathy, and autoimmune thrombocytopenia or neutropenia or pure red cell aplasias, while exemplary non-limiting examples of alloimmune diseases include allosensitization (see, for example, Blazar et al., 2015, Am. J. Transplant, 15(4):931-41) or xenosensitization from hematopoietic or solid organ transplantation, blood transfusions, pregnancy with fetal allosensitization, neonatal alloimmune thrombocytopenia, hemolytic disease of the newborn, sensitization to foreign antigens such as can occur with replacement of inherited or acquired deficiency disorders treated with enzyme or protein replacement therapy, blood products, and gene therapy. In some embodiments, the antigen characteristic of an autoimmune or inflammatory disorder is selected from a cell surface receptor, an ion channel-linked receptor, an enzyme-linked receptor, a G protein-coupled receptor, receptor tyrosine kinase, tyrosine kinase associated receptor, receptor-like tyrosine phosphatase, receptor serine/ threonine kinase, receptor guanylyl cyclase, or histidine kinase associated receptor.

**[0410]** In some embodiments, an antigen binding domain of a CAR binds to a ligand expressed on B cells, plasma cells, or plasmablasts. In some embodiments, an antigen binding domain of a CAR binds to CD10, CD19, CD20, CD22, CD24, CD27, CD38, CD45R, CD138, CD319, BCMA, CD28, TNF, interferon receptors, GM-CSF, ZAP-70, LFA-1, CD3 gamma, CD5 or CD2. See, *e.g.*, US 2003/0077249; WO 2017/058753; WO 2017/058850, the contents of which are herein incorporated by reference.

**[0411]** In some embodiments, the antigen binding domain targets an antigen characteristic of senescent cells, *e.g.*, urokinase-type plasminogen activator receptor (uPAR). In some embodiments, the ABD binds an antigen associated with a senescent cell. In some instances, the antigen is expressed by a senescent cell. In some embodiments, the CAR may be used for treatment or prophylaxis of disorders characterized by the aberrant accumulation of senescent cells, *e.g.*, liver and lung fibrosis, atherosclerosis, diabetes and osteoarthritis.

**[0412]** In some embodiments, the antigen binding domain targets an antigen characteristic of an infectious disease. In some embodiments, the ABD binds an antigen associated with an infectious disease. In some instances, the antigen is expressed by a cell affected by an infectious disease. In some embodiments, wherein the infectious disease is selected from HIV, hepatitis B virus, hepatitis C virus, Human herpes virus, Human herpes virus 8 (HHV-8, Kaposi sarcoma-associated herpes virus (KSHV)), Human T-lymphotrophic virus-1 (HTLV-1), Merkel cell polyomavirus (MCV), Simian virus 40 (SV40), Epstein-Barr virus, CMV, human papillomavirus. In some embodiments, the antigen characteristic of an infectious disease is selected from a cell surface receptor, an ion channel-linked receptor, an enzyme-linked receptor, a G protein-coupled receptor, receptor tyrosine kinase, tyrosine kinase associated receptor, receptor-like tyrosine phosphatase, receptor serine/ threonine kinase, receptor guanylyl cyclase, histidine kinase associated receptor, HIV Env, gpl20, or CD4-induced epitope on HIV-1 Env.

**[0413]** In some embodiments, an antigen binding domain binds to a cell surface antigen of a cell. In some embodiments, a cell surface antigen is characteristic of (*e.g.*, expressed by) a particular or specific cell type. In some embodiments, a cell surface antigen is characteristic of more than one type of cell.

**[0414]** In some embodiments, a CAR antigen binding domain binds a cell surface antigen characteristic of a T cell, such as a cell surface antigen on a T cell. In some embodiments, an antigen characteristic of a T cell may be a cell surface receptor, a membrane transport protein (*e.g.*, an active or passive transport protein such as, for example, an ion channel protein, a pore-forming protein, etc.), a transmembrane receptor, a membrane enzyme, and/or a cell adhesion protein characteristic of a T cell. In some embodiments, an antigen characteristic of a T cell may be a G protein-coupled receptor, receptor tyrosine kinase, tyrosine kinase associated receptor, receptor-like tyrosine phosphatase, receptor serine/ threonine kinase, receptor guanylyl cyclase, or histidine kinase associated receptor.

**[0415]** In some embodiments, the CAR comprises an extracellular antigen binding domain (e.g., antibody or antibody fragment, such as an scFv) that binds to an antigen (e.g. tumor antigen), a spacer (e.g. containing a hinge domain, such as any as described herein), a transmembrane domain (e.g. any as described herein), and an intracellular signaling domain (e.g. any intracellular signaling domain, such as a primary signaling domain or costimulatory signaling domain as described herein). In some embodiments, the intracellular signaling domain is or includes a primary cytoplasmic signaling domain. In some embodiments, the intracellular signaling domain additionally includes an intracellular signaling domain of a costimulatory molecule (e.g., a costimulatory domain).

**[0416]** In some embodiments, the antigen receptor further includes a marker and/or cells expressing the CAR or other antigen receptor further includes a surrogate marker, such as a cell surface marker, which may be used to confirm transduction or engineering of the cell to express the receptor. In some aspects, the marker includes all or part (e.g., truncated form) of CD34, a NGFR, or epidermal growth factor receptor, such as truncated version of such a cell surface receptor (e.g., tEGFR). In some embodiments, the nucleic acid encoding the marker is operably linked to a polynucleotide encoding for a linker sequence, such as a cleavable linker sequence, e.g., T2A. For example, a marker, and optionally a linker sequence, can be any as disclosed in published patent application No. WO2014031687. For example, the marker can be a truncated EGFR (tEGFR) that is, optionally, linked to a linker sequence, such as a T2A cleavable linker sequence.

**[0417]** In some embodiments, the marker is a molecule, e.g., cell surface protein, not naturally found on T cells or not naturally found on the surface of T cells, or a portion thereof. In some embodiments, the molecule is a non-self molecule, e.g., non-self protein, i.e., one that is not recognized as "self' by the immune system of the host into which the cells will be adoptively transferred.

**[0418]** In some embodiments, the marker serves no therapeutic function and/or produces no effect other than to be used as a marker for genetic engineering, e.g., for selecting cells successfully engineered. In other embodiments, the marker may be a therapeutic molecule or molecule otherwise exerting some

desired effect, such as a ligand for a cell to be encountered in vivo, such as a costimulatory or immune checkpoint molecule to enhance and/or dampen responses of the cells upon adoptive transfer and encounter with ligand.

**[0419]** In some cases, CARs are referred to as first, second, and/or third generation CARs. In some aspects, a first generation CAR is one that solely provides a CD3-chain induced signal upon antigen binding; in some aspects, a second-generation CARs is one that provides such a signal and costimulatory signal, such as one including an intracellular signaling domain from a costimulatory receptor such as CD28 or CD 137; in some aspects, a third generation CAR is one that includes multiple costimulatory domains of different costimulatory receptors.

**[0420]** For example, in some embodiments, the CAR contains an antibody, e.g., an antibody fragment, a transmembrane domain that is or contains a transmembrane portion of CD28 or a functional variant thereof, and an intracellular signaling domain containing a signaling portion of CD28 or functional variant thereof and a signaling portion of CD3 zeta or functional variant thereof. In some embodiments, the CAR contains an antibody, e.g., antibody fragment, a transmembrane domain that is or contains a transmembrane portion of CD28 or a functional variant thereof, and an intracellular signaling domain containing a signaling portion of a 4- IBB or functional variant thereof and a signaling portion of CD3 zeta or functional variant thereof. In some such embodiments, the receptor further includes a spacer containing a portion of an Ig molecule, such as a human Ig molecule, such as an Ig hinge, e.g. an IgG4 hinge, such as a hinge -only spacer.

**[0421]** In some aspects, the spacer contains only a hinge region of an IgG, such as only a hinge of IgG4 or IgGl. In other embodiments, the spacer is or contains an Ig hinge, e.g., an IgG4-derived hinge, optionally linked to a CH2 and/or CH3 domains. In some embodiments, the spacer is an Ig hinge, e.g., an IgG4 hinge, linked to CH2 and CH3 domains. In some embodiments, the spacer is an Ig hinge, e.g., an IgG4 hinge, linked to a CH3 domain only. In some embodiments, the spacer is or comprises a glycine-serine rich sequence or other flexible linker such as known flexible linkers.

**[0422]** For example, in some embodiments, the CAR includes an antibody such as an antibody fragment, including scFvs, a spacer, such as a spacer containing a portion of an immunoglobulin molecule, such as a hinge region and/or one or more constant regions of a heavy chain molecule, such as an Ig-hinge containing spacer, a transmembrane domain containing all or a portion of a CD28-derived transmembrane domain, a CD28 -derived intracellular signaling domain, and a CD3 zeta signaling domain. In some embodiments, the CAR includes an antibody or fragment, such as scFv, a spacer such as any of the Ig-hinge containing spacers, a CD28-derived transmembrane domain, a 4-1BB-derived intracellular signaling domain, and a CD3 zeta-derived signaling domain.

**[0423]** The recombinant receptors, such as CARs, expressed by the cells administered to the subject generally recognize or specifically bind to a molecule that is expressed in, associated with, and/or specific for the disease or condition or cells thereof being treated. Upon specific binding to the molecule,

e.g., antigen, the receptor generally delivers an immunostimulatory signal, such as an ITAM-transduced signal, into the cell, thereby promoting an immune response targeted to the disease or condition. For example, in some embodiments, the cells express a CAR that specifically binds to an antigen expressed by a cell or tissue of the disease or condition or associated with the disease or condition.

**[0424]** In some embodiments a lipid particle comprising a CAR or a nucleic acid encoding a CAR (e.g., a DNA, a gDNA, a cDNA, an RNA, a pre-MRNA, an mRNA, an miRNA, an siRNA, etc.) is delivered to a target cell. In some embodiments the target cell is an effector cell, e.g., a cell of the immune system that expresses one or more Fc receptors and mediates one or more effector functions. In some embodiments, a target cell may include, but may not be limited to, one or more of a monocyte, macrophage, neutrophil, dendritic cell, eosinophil, mast cell, platelet, large granular lymphocyte, Langerhans' cell, natural killer (NK) cell, T lymphocyte (e.g., T cell), a Gamma delta T cell, B lymphocyte (e.g., B cell) and may be from any organism including but not limited to humans, mice, rats, rabbits, and monkeys.

**[0425]** In certain embodiments, the exogenous agent is a CAR. CARs (also known as chimeric immunoreceptors, chimeric T cell receptors, or artificial T cell receptors) are receptor proteins that have been engineered to give host cells (e.g., T cells) the new ability to target a specific protein. The receptors are chimeric because they combine both antigen-binding and T cell activating functions into a single receptor. The provided particles may be used to express one or more CARs in a host cell (e.g., a T cell) for use in cell-based therapies against various target antigens. In these embodiments, the CAR may comprise an extracellular binding domain (also referred to as a "binder") that specifically binds a target antigen, a transmembrane domain, and an intracellular signaling domain. In certain embodiments, the CAR may further comprise one or more additional elements, including one or more signal peptides, one or more extracellular hinge domains, and/or one or more intracellular costimulatory domains. Domains may be directly adjacent to one another, or there may be one or more amino acids linking the domains. The nucleotide sequence encoding a CAR may be derived from a mammalian sequence, for example, a mouse sequence, a primate sequence, a human sequence, or combinations thereof. In the cases where the nucleotide sequence encoding a CAR is non-human, the sequence of the CAR may be humanized. The nucleotide sequence encoding a CAR may also be codon-optimized for expression in a mammalian cell, for example, a human cell. In any of these embodiments, the nucleotide sequence encoding a CAR may be at least 80% identical (*e.g.,* at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical) to any of the nucleotide sequences disclosed herein. The sequence variations may be due to codon-optimalization, humanization, restriction enzyme-based cloning scars, and/or additional amino acid residues linking the functional domains, etc.

**[0426]** In certain embodiments, the CAR may comprise a signal peptide at the N-terminus. Non-limiting examples of signal peptides include CD8α signal peptide, IgK signal peptide, and granulocyte-macrophage colony-stimulating factor receptor subunit alpha (GMCSFR-α, also known as colony

stimulating factor 2 receptor subunit alpha (CSF2RA)) signal peptide, and variants thereof, the amino acid sequences of which are provided in **Table 4** below.

**Table 4. Exemplary sequences of signal peptides**

| SEQ ID NO: | Sequence | Description |
|---|---|---|
| 286 | MALPVTALLLPLALLLHAARP | CD8α signal peptide |
| 287 | METDTLLLWVLLLWVPGSTG | IgK signal peptide |
| 288 | MLLLVTSLLLCELPHPAFLLIP | GMCSFR-α (CSF2RA) signal peptide |

**[0427]** In certain embodiments, the extracellular binding domain of the CAR may comprise one or more antibodies specific to one target antigen or multiple target antigens. The antibody may be an antibody fragment, for example, an scFv, or a single-domain antibody fragment, for example, a VHH. In certain embodiments, the scFv may comprise a heavy chain variable region ($V_H$) and a light chain variable region ($V_L$) of an antibody connected by a linker. The $V_H$ and the $V_L$ may be connected in either order, i.e., $V_H$-linker-$V_L$ or $V_L$-linker-$V_H$. Non-limiting examples of linkers include Whitlow linker, $(G_4S)_n$ (n can be a positive integer, *e.g.,* 1, 2, 3, 4, 5, 6, etc.) linker, and variants thereof. In certain embodiments, the antigen may be an antigen that is exclusively or preferentially expressed on tumor cells, or an antigen that is characteristic of an autoimmune or inflammatory disease. Exemplary target antigens include, but are not limited to, CD5, CD19, CD20, CD22, CD23, CD30, CD70, Kappa, Lambda, and B cell maturation agent (BCMA), G-protein coupled receptor family C group 5 member D (GPRC5D) (associated with leukemias); CS1/SLAMF7, CD38, CD138, GPRC5D, TACI, and BCMA (associated with myelomas); GD2, HER2, EGFR, EGFRvIII, B7H3, PSMA, PSCA, CAIX, CD171, CEA, CSPG4, EPHA2, FAP, FRα, IL-13Rα, Mesothelin, MUC1, MUC16, and ROR1 (associated with solid tumors). In any of these embodiments, the extracellular binding domain of the CAR can be codon-optimized for expression in a host cell or have variant sequences to increase functions of the extracellular binding domain.

**[0428]** In certain embodiments, the CAR may comprise a hinge domain, also referred to as a spacer. The terms "hinge" and "spacer" may be used interchangeably in the present disclosure. Non-limiting examples of hinge domains include CD8α hinge domain, CD28 hinge domain, IgG4 hinge domain, IgG4 hinge-CH2-CH3 domain, and variants thereof, the amino acid sequences of which are provided in **Table 5** below.

**Table 5. Exemplary sequences of hinge domains**

| SEQ ID NO: | Sequence | Description |
|---|---|---|
| 289 | TTTPAPRPPTPAPTIASQPLSLRPEACRPAAGGAVH TRGLDFACD | CD8α hinge domain |
| 290 | IEVMYPPPYLDNEKSNGTIIHVKGKHLCPSPLFPGP SKP | CD28 hinge domain |
| 291 | AAAIEVMYPPPYLDNEKSNGTIIHVKGKHLCPSPLF PGPSKP | CD28 hinge domain |
| 292 | ESKYGPPCPPCP | IgG4 hinge domain |
| 293 | ESKYGPPCPSCP | IgG4 hinge domain |
| 294 | ESKYGPPCPPCPAPEFLGGPSVFLFPPKPKDTLMISR TPEVTCVVVDVSQEDPEVQFNWYVDGVEVHNAK TKPREEQFNSTYRVVSVLTVLHQDWLNGKEYKCK VSNKGLPSSIEKTISKAKGQPREPQVYTLPPSQEEM TKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYK TTPPVLDSDGSFFLYSRLTVDKSRWQEGNVFSCSV MHEALHNHYTQKSLSLSLGK | IgG4 hinge-CH2-CH3 domain |

**[0429]** In certain embodiments, the transmembrane domain of the CAR may comprise a transmembrane region of the alpha, beta, or zeta chain of a T cell receptor, CD28, CD3ε, CD45, CD4, CD5, CD8, CD9, CD16, CD22, CD33, CD37, CD64, CD80, CD86, CD134, CD137, CD154, or a functional variant thereof, including the human versions of each of these sequences. In other embodiments, the transmembrane domain may comprise a transmembrane region of CD8α, CD8β, 4-1BB/CD137, CD28, CD34, CD4, FcεRIγ, CD16, OX40/CD134, CD3ζ, CD3ε, CD3γ, CD3δ, TCRα, TCRβ, TCRζ, CD32, CD64, CD64, CD45, CD5, CD9, CD22, CD37, CD80, CD86, CD40, CD40L/CD154, VEGFR2, FAS, and FGFR2B, or a functional variant thereof, including the human versions of each of these sequences. **Table 6** provides the amino acid sequences of a few exemplary transmembrane domains.

**Table 6. Exemplary sequences of transmembrane domains**

| SEQ ID NO: | Sequence | Description |
|---|---|---|
| 295 | IYIWAPLAGTCGVLLLSLVITLYC | CD8α transmembrane domain |
| 296 | FWVLVVVGGVLACYSLLVTVAFIIFWV | CD28 transmembrane domain |
| 297 | MFWVLVVVGGVLACYSLLVTVAFIIFWV | CD28 transmembrane domain |

**[0430]** In certain embodiments, the intracellular signaling domain and/or intracellular costimulatory domain of the CAR may comprise one or more signaling domains selected from B7-1/CD80, B7-2/CD86, B7-H1/PD-L1, B7-H2, B7-H3, B7-H4, B7-H6, B7-H7, BTLA/CD272, CD28, CTLA-4, Gi24/VISTA/B7-H5, ICOS/CD278, PD-1, PD-L2/B7-DC, PDCD6, 4-1BB/TNFSF9/CD137, 4-1BB

Ligand/TNFSF9, BAFF/BLyS/TNFSF13B, BAFF R/TNFRSF13C, CD27/TNFRSF7, CD27 Ligand/TNFSF7, CD30/TNFRSF8, CD30 Ligand/TNFSF8, CD40/TNFRSF5, CD40/TNFSF5, CD40 Ligand/TNFSF5, DR3/TNFRSF25, GITR/TNFRSF18, GITR Ligand/TNFSF18, HVEM/TNFRSF14, LIGHT/TNFSF14, Lymphotoxin-alpha/TNFβ, OX40/TNFRSF4, OX40 Ligand/TNFSF4, RELT/TNFRSF19L, TACI/TNFRSF13B, TL1A/TNFSF15, TNFα, TNF RII/TNFRSF1B, 2B4/CD244/SLAMF4, BLAME/SLAMF8, CD2, CD2F-10/SLAMF9, CD48/SLAMF2, CD58/LFA-3, CD84/SLAMF5, CD229/SLAMF3, CRACC/SLAMF7, NTB-A/SLAMF6, SLAM/CD150, CD2, CD7, CD53, CD82/Kai-1, CD90/Thy1, CD96, CD160, CD200, CD300a/LMIR1, HLA Class I, HLA-DR, Ikaros, Integrin alpha 4/CD49d, Integrin alpha 4 beta 1, Integrin alpha 4 beta 7/LPAM-1, LAG-3, TCL1A, TCL1B, CRTAM, DAP12, Dectin-1/CLEC7A, DPPIV/CD26, EphB6, TIM-1/KIM-1/HAVCR, TIM-4, TSLP, TSLP R, lymphocyte function associated antigen-1 (LFA-1), NKG2C, CD3ζ, an immunoreceptor tyrosine-based activation motif (ITAM), CD27, CD28, 4-1BB, CD134/OX40, CD30, CD40, PD-1, ICOS, lymphocyte function-associated antigen-1 (LFA-1), CD2, CD7, LIGHT, NKG2C, B7-H3, a ligand that specifically binds with CD83, and a functional variant thereof including the human versions of each of these sequences. In some embodiments, the intracellular signaling domain and/or intracellular costimulatory domain comprises one or more signaling domains selected from a CD3ζ domain, an ITAM, a CD28 domain, 4-1BB domain, or a functional variant thereof. **Table 7** provides the amino acid sequences of a few exemplary intracellular costimulatory and/or signaling domains. In certain embodiments, as in the case of tisagenlecleucel as described below, the CD3ζ signaling domain of SEQ ID NO:347 may have a mutation, *e.g.,* a glutamine (Q) to lysine (K) mutation, at amino acid position 14 (see SEQ ID NO:300).

**Table 7. Exemplary sequences of intracellular costimulatory and/or signaling domains**

| SEQ ID NO: | Sequence | Description |
|---|---|---|
| 298 | KRGRKKLLYIFKQPFMRPVQTTQEEDGCSCR FPEEEEGGCEL | 4-1BB costimulatory domain |
| 299 | RSKRSRLLHSDYMNMTPRRPGPTRKHYQPY APPRDFAAYRS | CD28 costimulatory domain |
| 300 | RVKFSRSADAPAYQQGQNQLYNELNLGRRE EYDVLDKRRGRDPEMGGKPRRKNPQEGLYN ELQKDKMAEAYSEIGMKGERRRGKGHDGLY QGLSTATKDTYDALHMQALPPR | CD3ζ signaling domain |
| 301 | RVKFSRSADAPAYKQGQNQLYNELNLGRRE EYDVLDKRRGRDPEMGGKPRRKNPQEGLYN ELQKDKMAEAYSEIGMKGERRRGKGHDGLY QGLSTATKDTYDALHMQALPPR | CD3ζ signaling domain (with Q to K mutation at position 14) |

*A) CD19 CAR*

**[0431]** In some embodiments, the CAR is a CD19 CAR ("CD19-CAR"). In some embodiments, the CD19 CAR may comprise a signal peptide, an extracellular binding domain that specifically binds CD19, a hinge domain, a transmembrane domain, an intracellular costimulatory domain, and/or an intracellular signaling domain in tandem.

**[0432]** In some embodiments, the signal peptide of the CD19 CAR comprises a CD8α signal peptide. In some embodiments, the CD8α signal peptide comprises or consists of an amino acid sequence set forth in SEQ ID NO:286 or an amino acid sequence that is at least 80% identical (*e.g.,* at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical) to the amino acid sequence set forth in of SEQ ID NO:286. In some embodiments, the signal peptide comprises an IgK signal peptide. In some embodiments, the IgK signal peptide comprises or consists of an amino acid sequence set forth in SEQ ID NO:287 or an amino acid sequence that is at least 80% identical (*e.g.,* at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical) to the amino acid sequence set forth in of SEQ ID NO:287. In some embodiments, the signal peptide comprises a GMCSFR-α or CSF2RA signal peptide. In some embodiments, the GMCSFR-α or CSF2RA signal peptide comprises or consists of an amino acid sequence set forth in SEQ ID NO:288 or an amino acid sequence that is at least 80% identical (*e.g.,* at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical) to the amino acid sequence set forth in of SEQ ID NO:288.

**[0433]** In some embodiments, the extracellular binding domain of the CD19 CAR is specific to CD19, for example, human CD19. The extracellular binding domain of the CD19 CAR can be codon-optimized for expression in a host cell or to have variant sequences to increase functions of the extracellular binding domain. In some embodiments, the extracellular binding domain comprises an immunogenically active portion of an immunoglobulin molecule, for example, an scFv.

**[0434]** In some embodiments, the extracellular binding domain of the CD19 CAR comprises an scFv derived from the FMC63 monoclonal antibody (FMC63), which comprises the heavy chain variable region ($V_H$) and the light chain variable region ($V_L$) of FMC63 connected by a linker. FMC63 and the derived scFv have been described in Nicholson et al., Mol. Immun. 34(16-17):1157-1165 (1997) and PCT Application Publication No. WO2018/213337, the entire contents of each of which are incorporated by reference herein. In some embodiments, the amino acid sequences of the entire FMC63-derived scFv (also referred to as FMC63 scFv) and its different portions are provided in **Table 8** below. In some embodiments, the CD19-specific scFv comprises or consists of an amino acid sequence set forth in SEQ ID NO:302, 303, or 308, or an amino acid sequence that is at least 80% identical (*e.g.,* at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical) to the amino acid sequence set forth in SEQ ID NO:302, 303, or 308. In some embodiments,

the CD19-specific scFv may comprise one or more CDRs having amino acid sequences set forth in SEQ ID NOs: 304-306 and 309-311. In some embodiments, the CD19-specific scFv may comprise a light chain with one or more CDRs having amino acid sequences set forth in SEQ ID NOs: 304-306. In some embodiments, the CD19-specific scFv may comprise a heavy chain with one or more CDRs having amino acid sequences set forth in SEQ ID NOs: 309-311. In any of these embodiments, the CD19-specific scFv may comprise one or more CDRs comprising one or more amino acid substitutions, or comprising a sequence that is at least 80% identical (*e.g.,* at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical), to any of the sequences identified. In some embodiments, the extracellular binding domain of the CD19 CAR comprises or consists of the one or more CDRs as described herein.

**[0435]** In some embodiments, the linker linking the $V_H$ and the $V_L$ portions of the scFv is a Whitlow linker having an amino acid sequence set forth in SEQ ID NO:307. In some embodiments, the Whitlow linker may be replaced by a different linker, for example, a 3x$G_4S$ linker having an amino acid sequence set forth in SEQ ID NO:313, which gives rise to a different FMC63-derived scFv having an amino acid sequence set forth in SEQ ID NO:312. In certain of these embodiments, the CD19-specific scFv comprises or consists of an amino acid sequence set forth in SEQ ID NO:312 or an amino acid sequence that is at least 80% identical (*e.g.,* at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical) to the amino acid sequence set forth in of SEQ ID NO:312.

**Table 8. Exemplary sequences of anti-CD19 scFv and components**

| SEQ ID NO: | Amino Acid Sequence | Description |
|---|---|---|
| 302 | DIQMTQTTSSLSASLGDRVTISCRASQDI SKYLNWYQQKPDGTVKLLIYHTSRLHS GVPSRFSGSGSGTDYSLTISNLEQEDIAT YFCQQGNTLPYTFGGGTKLEITGSTSGS GKPGSGEGSTKGEVKLQESGPGLVAPS QSLSVTCTVSGVSLPDYGVSWIRQPPRK GLEWLGVIWGSETTYYNSALKSRLTIIK DNSKSQVFLKMNSLQTDDTAIYYCAKH YYYGGSYAMDYWGQGTSVTVSS | Anti-CD19 FMC63 scFv entire sequence, with Whitlow linker |
| 303 | DIQMTQTTSSLSASLGDRVTISCRASQDI SKYLNWYQQKPDGTVKLLIYHTSRLHS GVPSRFSGSGSGTDYSLTISNLEQEDIAT YFCQQGNTLPYTFGGGTKLEIT | Anti-CD19 FMC63 scFv light chain variable region |
| 304 | QDISKY | Anti-CD19 FMC63 scFv light chain CDR1 |
| 305 | HTS | Anti-CD19 FMC63 scFv light chain CDR2 |
| 306 | QQGNTLPYT | Anti-CD19 FMC63 scFv light chain CDR3 |

| SEQ ID NO: | Amino Acid Sequence | Description |
|---|---|---|
| 307 | GSTSGSGKPGSGEGSTKG | Whitlow linker |
| 308 | EVKLQESGPGLVAPSQSLSVTCTVSGVS LPDYGVSWIRQPPRKGLEWLGVIWGSE TTYYNSALKSRLTIIKDNSKSQVFLKMN SLQTDDTAIYYCAKHYYYGGSYAMDY WGQGTSVTVSS | Anti-CD19 FMC63 scFv heavy chain variable region |
| 309 | GVSLPDYG | Anti-CD19 FMC63 scFv heavy chain CDR1 |
| 310 | IWGSETT | Anti-CD19 FMC63 scFv heavy chain CDR2 |
| 311 | AKHYYYGGSYAMDY | Anti-CD19 FMC63 scFv heavy chain CDR3 |
| 312 | DIQMTQTTSSLSASLGDRVTISCRASQDI SKYLNWYQQKPDGTVKLLIYHTSRLHS GVPSRFSGSGSGTDYSLTISNLEQEDIAT YFCQQGNTLPYTFGGGTKLEITGGGGS GGGGSGGGGSEVKLQESGPGLVAPSQS LSVTCTVSGVSLPDYGVSWIRQPPRKGL EWLGVIWGSETTYYNSALKSRLTIIKDN SKSQVFLKMNSLQTDDTAIYYCAKHYY YGGSYAMDYWGQGTSVTVSS | Anti-CD19 FMC63 scFv entire sequence, with 3x$G_4S$ linker |
| 313 | GGGGSGGGGSGGGGS | 3x$G_4S$ linker |

**[0436]** In some embodiments, the extracellular binding domain of the CD19 CAR is derived from an antibody specific to CD19, including, for example, SJ25C1 (Bejcek et al., Cancer Res. 55:2346-2351 (1995)), HD37 (Pezutto et al., J. Immunol. 138(9):2793-2799 (1987)), 4G7 (Meeker et al., Hybridoma 3:305-320 (1984)), B43 (Bejcek (1995)), BLY3 (Bejcek (1995)), B4 (Freedman et al., 70:418-427 (1987)), B4 HB12b (Kansas & Tedder, J. Immunol. 147:4094-4102 (1991); Yazawa et al., Proc. Natl. Acad. Sci. USA 102:15178-15183 (2005); Herbst et al., J. Pharmacol. Exp. Ther. 335:213-222 (2010)), BU12 (Callard et al., J. Immunology, 148(10): 2983-2987 (1992)), and CLB-CD19 (De Rie Cell. Immunol. 118:368-381(1989)). In any of these embodiments, the extracellular binding domain of the CD19 CAR can comprise or consist of the $V_H$, the $V_L$, and/or one or more CDRs of any of the antibodies.

**[0437]** In some embodiments, the hinge domain of the CD19 CAR comprises a CD8α hinge domain, for example, a human CD8α hinge domain. In some embodiments, the CD8α hinge domain comprises or consists of an amino acid sequence set forth in SEQ ID NO:289 or an amino acid sequence that is at least 80% identical (*e.g.*, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical) to the amino acid sequence set forth in of SEQ ID NO:289. In some embodiments, the hinge domain comprises a CD28 hinge domain, for example, a human CD28 hinge domain. In some embodiments, the CD28 hinge domain comprises or consists of an amino acid sequence set forth in SEQ ID NO:290 or an amino acid sequence that is at least 80% identical (*e.g.*, at

least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical) to the amino acid sequence set forth in of SEQ ID NO:290. In some embodiments, the hinge domain comprises an IgG4 hinge domain, for example, a human IgG4 hinge domain. In some embodiments, the IgG4 hinge domain comprises or consists of an amino acid sequence set forth in SEQ ID NO:292 or SEQ ID NO:293, or an amino acid sequence that is at least 80% identical (*e.g.,* at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical) to the amino acid sequence set forth in of SEQ ID NO:292 or SEQ ID NO:293. In some embodiments, the hinge domain comprises a IgG4 hinge-Ch2-Ch3 domain, for example, a human IgG4 hinge-Ch2-Ch3 domain. In some embodiments, the IgG4 hinge-Ch2-Ch3 domain comprises or consists of an amino acid sequence set forth in SEQ ID NO:293 or an amino acid sequence that is at least 80% identical (*e.g.,* at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical) to the amino acid sequence set forth in of SEQ ID NO:293.

**[0438]** In some embodiments, the transmembrane domain of the CD19 CAR comprises a CD8α transmembrane domain, for example, a human CD8α transmembrane domain. In some embodiments, the CD8α transmembrane domain comprises or consists of an amino acid sequence set forth in SEQ ID NO:295 or an amino acid sequence that is at least 80% identical (*e.g.,* at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical) to the amino acid sequence set forth in SEQ ID NO:295. In some embodiments, the transmembrane domain comprises a CD28 transmembrane domain, for example, a human CD28 transmembrane domain. In some embodiments, the CD28 transmembrane domain comprises or consists of an amino acid sequence set forth in SEQ ID NO:296 or 297 or an amino acid sequence that is at least 80% identical (*e.g.,* at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical) to the amino acid sequence set forth in SEQ ID NO:296 or 297.

**[0439]** In some embodiments, the intracellular costimulatory domain of the CD19 CAR comprises a 4-1BB costimulatory domain. 4-1BB, also known as CD137, transmits a potent costimulatory signal to T cells, promoting differentiation and enhancing long-term survival of T lymphocytes. In some embodiments, the 4-1BB costimulatory domain is human. In some embodiments, the 4-1BB costimulatory domain comprises or consists of an amino acid sequence set forth in SEQ ID NO:298 or an amino acid sequence that is at least 80% identical (*e.g.,* at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical) to the amino acid sequence set forth in SEQ ID NO:298. In some embodiments, the intracellular costimulatory domain comprises a CD28 costimulatory domain. CD28 is another co-stimulatory molecule on T cells. In some embodiments, the CD28 costimulatory domain is human. In some embodiments, the CD28 costimulatory domain comprises or consists of an amino acid sequence set forth in SEQ ID NO:299 or an amino acid sequence that is at least 80% identical (*e.g.,* at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical) to the amino acid sequence set forth in

SEQ ID NO:299. In some embodiments, the intracellular costimulatory domain of the CD19 CAR comprises a 4-1BB costimulatory domain and a CD28 costimulatory domain as described.

**[0440]** In some embodiments, the intracellular signaling domain of the CD19 CAR comprises a CD3 zeta (ζ) signaling domain. CD3 zeta associates with T cell receptors (TCRs) to produce a signal and contains immunoreceptor tyrosine-based activation motifs (ITAMs). The CD3 zeta signaling domain refers to amino acid residues from the cytoplasmic domain of the zeta chain that are sufficient to functionally transmit an initial signal necessary for T cell activation. In some embodiments, the CD3 zeta signaling domain is human. In some embodiments, the CD3 zeta signaling domain comprises or consists of an amino acid sequence set forth in SEQ ID NO:300 or an amino acid sequence that is at least 80% identical (*e.g.,* at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical) to the amino acid sequence set forth in SEQ ID NO:300.

**[0441]** In some embodiments, the payload agent is a CD19 CAR, including, for example, a CD19 CAR comprising the CD19-specific scFv having sequences set forth in SEQ ID NO:302 or SEQ ID NO:312, the CD8α hinge domain of SEQ ID NO:289, the CD8α transmembrane domain of SEQ ID NO:294, the 4-1BB costimulatory domain of SEQ ID NO:298, the CD3ζ signaling domain of SEQ ID NO:300, and/or variants (i.e., having a sequence that is at least 80% identical, for example, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99 identical to the disclosed sequence) thereof. In any of these embodiments, the CD19 CAR may additionally comprise a signal peptide (*e.g.,* a CD8α signal peptide) as described.

**[0442]** In some embodiments, the payload agent is a CD19 CAR, including, for example, a CD19 CAR comprising the CD19-specific scFv having sequences set forth in SEQ ID NO:302 or SEQ ID NO:312, the IgG4 hinge domain of SEQ ID NO:292 or SEQ ID NO:293, the CD28 transmembrane domain of SEQ ID NO:296, the 4-1BB costimulatory domain of SEQ ID NO:298, the CD3ζ signaling domain of SEQ ID NO:300, and/or variants (i.e., having a sequence that is at least 80% identical, for example, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99 identical to the disclosed sequence) thereof. In any of these embodiments, the CD19 CAR may additionally comprise a signal peptide (*e.g.,* a CD8α signal peptide) as described.

**[0443]** In some embodiments, the payload agent is a CD19 CAR, including, for example, a CD19 CAR comprising the CD19-specific scFv having sequences set forth in SEQ ID NO:302 or SEQ ID NO:312, the CD28 hinge domain of SEQ ID NO:290, the CD28 transmembrane domain of SEQ ID NO:296, the CD28 costimulatory domain of SEQ ID NO:299, the CD3ζ signaling domain of SEQ ID NO:300, and/or variants (i.e., having a sequence that is at least 80% identical, for example, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99 identical to the disclosed sequence) thereof. In any of these embodiments, the CD19 CAR may additionally comprise a signal peptide (*e.g.,* a CD8α signal peptide) as described.

**[0444]** In some embodiments, the payload agent is a CD19 CAR as encoded by the sequence set forth in SEQ ID NO:314 or a sequence at least 80% identical (*e.g.,* at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical) to the nucleotide sequence set forth in SEQ ID NO:314 (see **Table 9**). The encoded CD19 CAR has a corresponding amino acid sequence set forth in SEQ ID NO:315 or is at least 80% identical (*e.g.,* at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical) to the amino acid sequence set forth in of SEQ ID NO:315, with the following components: CD8α signal peptide, FMC63 scFv ($V_L$-Whitlow linker-$V_H$), CD8α hinge domain, CD8α transmembrane domain, 4-1BB costimulatory domain, and CD3ζ signaling domain.

**[0445]** In some embodiments, the payload agent is a commercially available embodiment of a CD19 CAR. Non-limiting examples of commercially available embodiments of CD19 CARs include tisagenlecleucel, lisocabtagene maraleucel, axicabtagene ciloleucel, and brexucabtagene autoleucel.

**[0446]** In some embodiments, the CAR is tisagenlecleucel or portions thereof. Tisagenlecleucel comprises a CD19 CAR with the following components: CD8α signal peptide, FMC63 scFv ($V_L$-3x$G_4S$ linker-$V_H$), CD8α hinge domain, CD8α transmembrane domain, 4-1BB costimulatory domain, and CD3ζ signaling domain. The nucleotide and amino acid sequence of the CD19 CAR in tisagenlecleucel are provided in **Table 9**, with annotations of the sequences provided in **Table 10**.

**[0447]** In some embodiments, the CAR is lisocabtagene maraleucel or portions thereof. Lisocabtagene maraleucel comprises a CD19 CAR with the following components: GMCSFR-α or CSF2RA signal peptide, FMC63 scFv ($V_L$-Whitlow linker-$V_H$), IgG4 hinge domain, CD28 transmembrane domain, 4-1BB costimulatory domain, and CD3ζ signaling domain. The nucleotide and amino acid sequence of the CD19 CAR in lisocabtagene maraleucel are provided in **Table 9**, with annotations of the sequences provided in **Table 11**.

**[0448]** In some embodiments, the CAR is axicabtagene ciloleucel or portions thereof. Axicabtagene ciloleucel comprises a CD19 CAR with the following components: GMCSFR-α or CSF2RA signal peptide, FMC63 scFv ($V_L$-Whitlow linker-$V_H$), CD28 hinge domain, CD28 transmembrane domain, CD28 costimulatory domain, and CD3ζ signaling domain. The nucleotide and amino acid sequence of the CD19 CAR in axicabtagene ciloleucel are provided in **Table 9**, with annotations of the sequences provided in **Table 12**.

**[0449]** In some embodiments, the CAR is brexucabtagene autoleucel or portions thereof. Brexucabtagene autoleucel comprises a CD19 CAR with the following components: GMCSFR- α signal peptide, FMC63 scFv, CD28 hinge domain, CD28 transmembrane domain, CD28 costimulatory domain, and CD3ζ signaling domain.

**[0450]** In some embodiments, the CAR is encoded by the sequence set forth in SEQ ID NO: 316, 318, or 320, or a sequence at least 80% identical (*e.g.,* at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical) to the nucleotide sequence

set forth in SEQ ID NO: 316, 318, or 320. The encoded CD19 CAR has a corresponding amino acid sequence set forth in SEQ ID NO: 317, 319, or 321, respectively, or is at least 80% identical (*e.g.,* at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical) to the amino acid sequence set forth in of SEQ ID NO: 317, 319, or 321, respectively.

**Table 9. Exemplary sequences of CD19 CARs**

| SEQ ID NO: | Sequence | Description |
|---|---|---|
| 314 | atggccttaccagtgaccgccttgctcctgccgctggccttgctgctccac<br>gccgccaggccggacatccagatgacacagactacatcctccctgtctgc<br>ctctctgggagacagagtcaccatcagttgcagggcaagtcaggacatta<br>gtaaatatttaaattggtatcagcagaaaccagatggaactgttaaactcct<br>gatctaccatacatcaagattacactcaggagtcccatcaaggttcagtgg<br>cagtgggtctggaacagattattctctcaccattagcaacctggagcaaga<br>agatattgccacttacttttgccaacagggtaatacgcttccgtacacgttcg<br>gaggggggaccaagctggagatcacaggctccacctctggatccggca<br>agcccggatctggcgagggatccaccaagggcgaggtgaaactgcag<br>gagtcaggacctggcctggtggcgccctcacagagcctgtccgtcacat<br>gcactgtctcaggggtctcattacccgactatggtgtaagctggattcgcc<br>agcctccacgaaagggtctggagtggctgggagtaatatggggtagtga<br>aaccacatactataattcagctctcaaatccagactgaccatcatcaagga<br>caactccaagagccaagttttcttaaaaatgaacagtctgcaaactgatga<br>cacagccatttactactgtgccaaacattattactacggtggtagctatgcta<br>tggactactggggccaaggaacctcagtcaccgtctcctcaaccacgac<br>gccagcgccgcgaccaccaacaccggcgcccaccatcgcgtcgcagc<br>ccctgtccctgcgcccagaggcgtgccggccagcggcgggggggcgca<br>gtgcacacgagggggctggacttcgcctgtgatatctacatctgggcgcc<br>cttggccgggacttgtggggtccttctcctgtcactggttatcacccttact<br>gcaaacggggcagaaagaaactcctgtatatattcaaacaaccatttatga<br>gaccagtacaaactactcaagaggaagatggctgtagctgccgatttcca<br>gaagaagaagaaggaggatgtgaactgagagtgaagttcagcaggagc<br>gcagacgcccccgcgtaccagcagggccagaaccagctctataacgag<br>ctcaatctaggacgaagagaggagtacgatgttttggacaagagacgtgg<br>ccgggaccctgagatgggggggaaagccgagaaggaagaaccctcagg<br>aaggcctgtacaatgaactgcagaaagataagatggcggaggcctacag<br>tgagattgggatgaaaggcgagcgccggaggggcaaggggcacgatg<br>gcctttaccagggtctcagtacagccaccaaggacacctacgacgccctt<br>cacatgcaggccctgcccccctcgc | Exemplary CD19 CAR nucleotide sequence |
| 315 | MALPVTALLLPLALLLHAARPDIQMTQTTSSLS<br>ASLGDRVTISCRASQDISKYLNWYQQKPDGTV<br>KLLIYHTSRLHSGVPSRFSGSGSGTDYSLTISNL<br>EQEDIATYFCQQGNTLPYTFGGGTKLEITGSTS<br>GSGKPGSGEGSTKGEVKLQESGPGLVAPSQSLS<br>VTCTVSGVSLPDYGVSWIRQPPRKGLEWLGVI<br>WGSETTYYNSALKSRLTIIKDNSKSQVFLKMNS<br>LQTDDTAIYYCAKHYYYGGSYAMDYWGQGT<br>SVTVSSTTTPAPRPPTPAPTIASQPLSLRPEACRP<br>AAGGAVHTRGLDFACDIYIWAPLAGTCGVLLL<br>SLVITLYCKRGRKKLLYIFKQPFMRPVQTTQEE<br>DGCSCRFPEEEEGGCELRVKFSRSADAPAYQQ<br>GQNQLYNELNLGRREEYDVLDKRRGRDPEMG | Exemplary CD19 CAR amino acid sequence |

| SEQ ID NO: | Sequence | Description |
|---|---|---|
| | GKPRRKNPQEGLYNELQKDKMAEAYSEIGMK GERRRGKGHDGLYQGLSTATKDTYDALHMQA LPPR | |
| 316 | atggccttaccagtgaccgccttgctcctgccgctggccttgctgctccac gccgccaggccggacatccagatgacacagactacatcctccctgtctgc ctctctgggagacagagtcaccatcagttgcagggcaagtcaggacatta gtaaatatttaaattggtatcagcagaaaccagatggaactgttaaactcct gatctaccatacatcaagattacactcaggagtcccatcaaggttcagtgg cagtgggtctggaacagattattctctcaccattagcaacctggagcaaga agatattgccacttacttttgccaacagggtaatacgcttccgtacacgttcg gaggggggaccaagctggagatcacaggtggcggtggctcgggcggt ggtgggtcgggtggcggcggatctgaggtgaaactgcaggagtcagga cctggcctggtggcgccctcacagagcctgtccgtcacatgcactgtctc aggggtctcattacccgactatggtgtaagctggattcgccagcctccacg aaagggtctggagtggctgggagtaatatggggtagtgaaaccacatact ataattcagctctcaaatccagactgaccatcatcaaggacaactccaaga gccaagttttcttaaaaatgaacagtctgcaaactgatgacacagccattta ctactgtgccaaacattattactacggtggtagctatgctatggactactgg ggccaaggaacctcagtcaccgtctcctcaaccacgacgccagcgccg cgaccaccaacaccggcgcccaccatcgcgtcgcagcccctgtccctgc gcccagaggcgtgccggccagcggcgggggggcgcagtgcacacgag ggggctggacttcgcctgtgatatctacatctgggcgcccttggccggga cttgtggggtccttctcctgtcactggttatcacccttactgcaaacggggc agaaagaaactcctgtatatattcaaacaaccatttatgagaccagtacaaa ctactcaagaggaagatggctgtagctgccgatttccagaagaagaagaa ggaggatgtgaactgagagtgaagttcagcaggagcgcagacgcccc gcgtacaagcagggccagaaccagctctataacgagctcaatctaggac gaagagaggagtacgatgttttggacaagagacgtggccgggaccctga gatgggggggaaagccgagaaggaagaaccctcaggaaggcctgtaca atgaactgcagaaagataagatggcggaggcctacagtgagattgggat gaaaggcgagcgccggaggggcaaggggcacgatggccttaccagg gtctcagtacagccaccaaggacacctacgacgcccttcacatgcaggc cctgcccctcgc | Tisagenlecleucel CD19 CAR nucleotide sequence |
| 317 | MALPVTALLLPLALLLHAARPDIQMTQTTSSLS ASLGDRVTISCRASQDISKYLNWYQQKPDGTV KLLIYHTSRLHSGVPSRFSGSGSGTDYSLTISNL EQEDIATYFCQQGNTLPYTFGGGTKLEITGGGG SGGGGSGGGGSEVKLQESGPGLVAPSQSLSVT CTVSGVSLPDYGVSWIRQPPRKGLEWLGVIWG SETTYYNSALKSRLTIIKDNSKSQVFLKMNSLQ TDDTAIYYCAKHYYYGGSYAMDYWGQGTSV TVSSTTTPAPRPPTPAPTIASQPLSLRPEACRPA AGGAVHTRGLDFACDIYIWAPLAGTCGVLLLS LVITLYCKRGRKKLLYIFKQPFMRPVQTTQEED GCSCRFPEEEEGGCELRVKFSRSADAPAYKQG QNQLYNELNLGRREEYDVLDKRRGRDPEMGG KPRRKNPQEGLYNELQKDKMAEAYSEIGMKG ERRRGKGHDGLYQGLSTATKDTYDALHMQAL PPR | Tisagenlecleucel CD19 CAR amino acid sequence |

| SEQ ID NO: | Sequence | Description |
|---|---|---|
| 318 | atgctgctgctggtgaccagcctgctgctgtgcgagctgccccaccccgc<br>ctttctgctgatccccgacatccagatgacccagaccacctccagcctgag<br>cgccagcctgggcgaccgggtgaccatcagctgccgggccagccagg<br>acatcagcaagtacctgaactggtatcagcagaagcccgacggcaccgt<br>caagctgctgatctaccacaccagccggctgcacagcggcgtgcccagc<br>cggtttagcggcagcggctccggcaccgactacagcctgaccatctcca<br>acctggaacaggaagatatcgccacctacttttgccagcagggcaacaca<br>ctgccctacacctttggcggcggaacaaagctggaaatcaccggcagca<br>cctccggcagcggcaagcctggcagcggcgagggcagcaccaaggg<br>cgaggtgaagctgcaggaaagcggccctggcctggtggccccagcca<br>gagcctgagcgtgacctgcaccgtgagcggcgtgagcctgcccgacta<br>cggcgtgagctggatccggcagccccccaggaagggcctggaatggct<br>gggcgtgatctggggcagcgagaccacctactacaacagcgccctgaa<br>gagccggctgaccatcatcaaggacaacagcaagagccaggtgttcctg<br>aagatgaacagcctgcagaccgacgacaccgccatctactactgcgcca<br>agcactactactacggcggcagctacgccatggactactggggccaggg<br>caccagcgtgaccgtgagcagcgaatctaagtacggaccgccctgcccc<br>ccttgccctatgttctgggtgctggtggtggtcggaggcgtgctggcctgc<br>tacagcctgctggtcaccgtggccttcatcatcttttgggtgaaacggggc<br>agaaagaaactcctgtatatattcaaacaaccatttatgagaccagtacaaa<br>ctactcaagaggaagatggctgtagctgccgatttccagaagaagaagaa<br>ggaggatgtgaactgcgggtgaagttcagcagaagcgccgacgcccct<br>gcctaccagcagggccagaatcagctgtacaacgagctgaacctgggc<br>agaagggaagagtacgacgtcctggataagcggagaggccgggaccc<br>tgagatgggcggcaagcctcggcggaagaaccccaggaaggcctgta<br>taacgaactgcagaaagacaagatggccgaggcctacagcgagatcgg<br>catgaagggcgagcggaggcggggcaagggccacgacggcctgtatc<br>agggcctgtccaccgccaccaaggatacctacgacgccctgcacatgca<br>ggccctgcccccaagg | Lisocabtagene maraleucel CD19 CAR nucleotide sequence |
| 319 | MLLLVTSLLLCELPHPAFLLIPDIQMTQTTSSLS<br>ASLGDRVTISCRASQDISKYLNWYQQKPDGTV<br>KLLIYHTSRLHSGVPSRFSGSGSGTDYSLTISNL<br>EQEDIATYFCQQGNTLPYTFGGGTKLEITGSTS<br>GSGKPGSGEGSTKGEVKLQESGPGLVAPSQSLS<br>VTCTVSGVSLPDYGVSWIRQPPRKGLEWLGVI<br>WGSETTYYNSALKSRLTIIKDNSKSQVFLKMNS<br>LQTDDTAIYYCAKHYYYGGSYAMDYWGQGT<br>SVTVSSESKYGPPCPPCPMFWVLVVVGGVLAC<br>YSLLVTVAFIIFWVKRGRKKLLYIFKQPFMRPV<br>QTTQEEDGCSCRFPEEEEGGCELRVKFSRSADA<br>PAYQQGQNQLYNELNLGRREEYDVLDKRRGR<br>DPEMGGKPRRKNPQEGLYNELQKDKMAEAYS<br>EIGMKGERRRGKGHDGLYQGLSTATKDTYDA<br>LHMQALPPR | Lisocabtagene maraleucel CD19 CAR amino acid sequence |
| 320 | atgcttctcctggtgacaagccttctgctctgtgagttaccacacccagcatt<br>cctcctgatcccagacatccagatgacacagactacatcctccctgtctgc<br>ctctctgggagacagagtcaccatcagttgcagggcaagtcaggacatta<br>gtaaatatttaaattggtatcagcagaaaccagatggaactgttaaactcct<br>gatctaccatacatcaagattacactcaggagtcccatcaaggttcagtgg<br>cagtgggtctggaacagattattctctcaccattagcaacctggagcaaga<br>agatattgccacttacttttgccaacagggtaatacgcttccgtacacgttcg | Axicabtagene ciloleucel CD19 CAR nucleotide sequence |

| SEQ ID NO: | Sequence | Description |
|---|---|---|
| | gagggggactaagttggaaataacaggctccacctctggatccggcaa gcccggatctggcgagggatccaccaagggcgaggtgaaactgcagg agtcaggacctggcctggtggcgccctcacagagcctgtccgtcacatg cactgtctcaggggtctcattacccgactatggtgtaagctggattcgcca gcctccacgaaagggtctggagtggctgggagtaatatggggtagtgaa accacatactataattcagctctcaaatccagactgaccatcatcaaggac aactccaagagccaagttttcttaaaaatgaacagtctgcaaactgatgac acagccatttactactgtgccaaacattattactacggtggtagctatgctat ggactactggggtcaaggaacctcagtcaccgtctcctcagcggccgca attgaagttatgtatcctcctccttacctagacaatgagaagagcaatggaa ccattatccatgtgaaagggaaacacctttgtccaagtcccctatttcccgg accttctaagccctttgggtgctggtggtggttgggggagtcctggcttgc tatagcttgctagtaacagtggcctttattattttctgggtgaggagtaagag gagcaggctcctgcacagtgactacatgaacatgactccccgccgcccc gggcccacccgcaagcattaccagccctatgccccaccacgcgacttcg cagcctatcgctccagagtgaagttcagcaggagcgcagacgcccccg cgtaccagcagggccagaaccagctctataacgagctcaatctaggacg aagagaggagtacgatgttttggacaagagacgtggccgggaccctgag atgggggaaagccgagaaggaagaaccctcaggaaggcctgtacaat gaactgcagaaagataagatggcggaggcctacagtgagattgggatga aaggcgagcgccggaggggcaaggggcacgatggcctttaccagggt ctcagtacagccaccaaggacacctacgacgcccttcacatgcaggccc tgcccccctcgc | |
| 321 | MLLLVTSLLLCELPHPAFLLIPDIQMTQTTSSLS ASLGDRVTISCRASQDISKYLNWYQQKPDGTV KLLIYHTSRLHSGVPSRFSGSGSGTDYSLTISNL EQEDIATYFCQQGNTLPYTFGGGTKLEITGSTS GSGKPGSGEGSTKGEVKLQESGPGLVAPSQSLS VTCTVSGVSLPDYGVSWIRQPPRKGLEWLGVI WGSETTYYNSALKSRLTIIKDNSKSQVFLKMNS LQTDDTAIYYCAKHYYYGGSYAMDYWGQGT SVTVSSAAAIEVMYPPPYLDNEKSNGTIIHVKG KHLCPSPLFPGPSKPFWVLVVVGGVLACYSLL VTVAFIIFWVRSKRSRLLHSDYMNMTPRRPGPT RKHYQPYAPPRDFAAYRSRVKFSRSADAPAYQ QGQNQLYNELNLGRREEYDVLDKRRGRDPEM GGKPRRKNPQEGLYNELQKDKMAEAYSEIGM KGERRRGKGHDGLYQGLSTATKDTYDALHMQ ALPPR | Axicabtagene ciloleucel CD19 CAR amino acid sequence |

**Table 10. Annotation of tisagenlecleucel CD19 CAR sequences**

| Feature | Nucleotide Sequence Position | Amino Acid Sequence Position |
|---|---|---|
| CD8α signal peptide | 1-63 | 1-21 |
| FMC63 scFv ($V_L$-3x$G_4$S linker-$V_H$) | 64-789 | 22-263 |
| CD8α hinge domain | 790-924 | 264-308 |
| CD8α transmembrane domain | 925-996 | 309-332 |

| Feature | Nucleotide Sequence Position | Amino Acid Sequence Position |
|---|---|---|
| 4-1BB costimulatory domain | 997-1122 | 333-374 |
| CD3ζ signaling domain | 1123-1458 | 375-486 |

**Table 11. Annotation of lisocabtagene maraleucel CD19 CAR sequences**

| Feature | Nucleotide Sequence Position | Amino Acid Sequence Position |
|---|---|---|
| GMCSFR-α signal peptide | 1-66 | 1-22 |
| FMC63 scFv ($V_L$-Whitlow linker-$V_H$) | 67-801 | 23-267 |
| IgG4 hinge domain | 802-837 | 268-279 |
| CD28 transmembrane domain | 838-921 | 280-307 |
| 4-1BB costimulatory domain | 922-1047 | 308-349 |
| CD3ζ signaling domain | 1048-1383 | 350-461 |

**Table 12. Annotation of axicabtagene ciloleucel CD19 CAR sequences**

| Feature | Nucleotide Sequence Position | Amino Acid Sequence Position |
|---|---|---|
| CSF2RA signal peptide | 1-66 | 1-22 |
| FMC63 scFv ($V_L$-Whitlow linker-$V_H$) | 67-801 | 23-267 |
| CD28 hinge domain | 802-927 | 268-309 |
| CD28 transmembrane domain | 928-1008 | 310-336 |
| CD28 costimulatory domain | 1009-1131 | 337-377 |
| CD3 zeta signaling domain | 1132-1467 | 378-489 |

**[0451]** In some embodiments, the CAR is encoded by the sequence set forth in SEQ ID NO: 316, 318, or 320, or a sequence at least 80% identical (*e.g.*, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical) to the nucleotide sequence set forth in SEQ ID NO: 316, 318, or 320. The encoded CD19 CAR has a corresponding amino acid sequence set forth in SEQ ID NO: 317, 319, or 321, respectively, is at least 80% identical (*e.g.*, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical) to the amino acid sequence set forth in of SEQ ID NO: 317, 319, or 321, respectively.

*B) CD20 CAR*

**[0452]** In some embodiments, the CAR is a CD20 CAR ("CD20-CAR"). CD20 is an antigen found on the surface of B cells as early at the pro-B phase and progressively at increasing levels until B cell maturity, as well as on the cells of most B-cell neoplasms. CD20 positive cells are also sometimes found in cases of Hodgkin's disease, myeloma, and thymoma. In some embodiments, the CD20 CAR may

comprise a signal peptide, an extracellular binding domain that specifically binds CD20, a hinge domain, a transmembrane domain, an intracellular costimulatory domain, and/or an intracellular signaling domain in tandem.

**[0453]** In some embodiments, the signal peptide of the CD20 CAR comprises a CD8α signal peptide. In some embodiments, the CD8α signal peptide comprises or consists of an amino acid sequence set forth in SEQ ID NO:286 or an amino acid sequence that is at least 80% identical (*e.g.,* at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical) to the amino acid sequence set forth in of SEQ ID NO:286. In some embodiments, the signal peptide comprises an IgK signal peptide. In some embodiments, the IgK signal peptide comprises or consists of an amino acid sequence set forth in SEQ ID NO:287 or an amino acid sequence that is at least 80% identical (*e.g.,* at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical) to the amino acid sequence set forth in of SEQ ID NO:287. In some embodiments, the signal peptide comprises a GMCSFR-α or CSF2RA signal peptide. In some embodiments, the GMCSFR-α or CSF2RA signal peptide comprises or consists of an amino acid sequence set forth in SEQ ID NO:288 or an amino acid sequence that is at least 80% identical (*e.g.,* at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical) to the amino acid sequence set forth in of SEQ ID NO:288.

**[0454]** In some embodiments, the extracellular binding domain of the CD20 CAR is specific to CD20, for example, human CD20. The extracellular binding domain of the CD20 CAR can be codon-optimized for expression in a host cell or to have variant sequences to increase functions of the extracellular binding domain. In some embodiments, the extracellular binding domain comprises an immunogenically active portion of an immunoglobulin molecule, for example, an scFv.

**[0455]** In some embodiments, the extracellular binding domain of the CD20 CAR is derived from an antibody specific to CD20, including, for example, Leu16, IF5, 1.5.3, rituximab, obinutuzumab, ibritumomab, ofatumumab, tositumumab, odronextamab, veltuzumab, ublituximab, and ocrelizumab. In any of these embodiments, the extracellular binding domain of the CD20 CAR can comprise or consist of the $V_H$, the $V_L$, and/or one or more CDRs of any of the antibodies.

**[0456]** In some embodiments, the extracellular binding domain of the CD20 CAR comprises an scFv derived from the Leu16 monoclonal antibody, which comprises the heavy chain variable region ($V_H$) and the light chain variable region ($V_L$) of Leu16 connected by a linker. See Wu et al., Protein Engineering. 14(12):1025-1033 (2001). In some embodiments, the linker is a 3x$G_4S$ linker. In other embodiments, the linker is a Whitlow linker as described herein. In some embodiments, the amino acid sequences of different portions of the entire Leu16-derived scFv (also referred to as Leu16 scFv) and its different portions are provided in **Table 13** below. In some embodiments, the CD20-specific scFv comprises or consists of an amino acid sequence set forth in SEQ ID NO:322, 323, or 327, or an amino acid sequence that is at least 80% identical (*e.g.,* at least 80%, at least 85%, at least 90%, at least 95%, at

least 96%, at least 97%, at least 98%, at least 99%, or 100% identical) to the amino acid sequence set forth in of SEQ ID NO: 322, 323, or 327. In some embodiments, the CD20-specific scFv may comprise one or more CDRs having amino acid sequences set forth in SEQ ID NOs: 324-326, 328, 329, and 330. In some embodiments, the CD20-specific scFv may comprise a light chain with one or more CDRs having amino acid sequences set forth in SEQ ID NOs: 324-326. In some embodiments, the CD20-specific scFv may comprise a heavy chain with one or more CDRs having amino acid sequences set forth in SEQ ID NOs: 328, 329, and 330. In any of these embodiments, the CD20-specific scFv may comprise one or more CDRs comprising one or more amino acid substitutions, or comprising a sequence that is at least 80% identical (*e.g.,* at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical), to any of the sequences identified. In some embodiments, the extracellular binding domain of the CD20 CAR comprises or consists of the one or more CDRs as described herein.

**Table 13. Exemplary sequences of anti-CD20 scFv and components**

| SEQ ID NO: | Amino Acid Sequence | Description |
|---|---|---|
| 322 | DIVLTQSPAILSASPGEKVTMTCRASSS VNYMDWYQKKPGSSPKPWIYATSNLA SGVPARFSGSGSGTSYSLTISRVEAEDA ATYYCQQWSFNPPTFGGGTKLEIKGSTS GSGKPGSGEGSTKGEVQLQQSGAELVK PGASVKMSCKASGYTFTSYNMHWVKQ TPGQGLEWIGAIYPGNGDTSYNQKFKG KATLTADKSSSTAYMQLSSLTSEDSAD YYCARSNYYGSSYWFFDVWGAGTTVT VSS | Anti-CD20 Leu16 scFv entire sequence, with Whitlow linker |
| 323 | DIVLTQSPAILSASPGEKVTMTCRASSS VNYMDWYQKKPGSSPKPWIYATSNLA SGVPARFSGSGSGTSYSLTISRVEAEDA ATYYCQQWSFNPPTFGGGTKLEIK | Anti-CD20 Leu16 scFv light chain variable region |
| 324 | RASSSVNYMD | Anti-CD20 Leu16 scFv light chain CDR1 |
| 325 | ATSNLAS | Anti-CD20 Leu16 scFv light chain CDR2 |
| 326 | QQWSFNPPT | Anti-CD20 Leu16 scFv light chain CDR3 |
| 327 | EVQLQQSGAELVKPGASVKMSCKASG YTFTSYNMHWVKQTPGQGLEWIGAIYP GNGDTSYNQKFKGKATLTADKSSSTAY MQLSSLTSEDSADYYCARSNYYGSSYW FFDVWGAGTTVTVSS | Anti-CD20 Leu16 scFv heavy chain |
| 328 | SYNMH | Anti-CD20 Leu16 scFv heavy chain CDR1 |

| SEQ ID NO: | Amino Acid Sequence | Description |
|---|---|---|
| 329 | AIYPGNGDTSYNQKFKG | Anti-CD20 Leu16 scFv heavy chain CDR2 |
| 330 | SNYYGSSYWFFDV | Anti-CD20 Leu16 scFv heavy chain CDR3 |

**[0457]** In some embodiments, the hinge domain of the CD20 CAR comprises a CD8α hinge domain, for example, a human CD8α hinge domain. In some embodiments, the CD8α hinge domain comprises or consists of an amino acid sequence set forth in SEQ ID NO:289 or an amino acid sequence that is at least 80% identical (*e.g.,* at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical) to the amino acid sequence set forth in of SEQ ID NO:289. In some embodiments, the hinge domain comprises a CD28 hinge domain, for example, a human CD28 hinge domain. In some embodiments, the CD28 hinge domain comprises or consists of an amino acid sequence set forth in SEQ ID NO:290 or an amino acid sequence that is at least 80% identical (*e.g.,* at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical) to the amino acid sequence set forth in of SEQ ID NO:290. In some embodiments, the hinge domain comprises an IgG4 hinge domain, for example, a human IgG4 hinge domain. In some embodiments, the IgG4 hinge domain comprises or consists of an amino acid sequence set forth in SEQ ID NO:292 or SEQ ID NO:293, or an amino acid sequence that is at least 80% identical (*e.g.,* at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical) to the amino acid sequence set forth in of SEQ ID NO:292 or SEQ ID NO:293. In some embodiments, the hinge domain comprises a IgG4 hinge-Ch2-Ch3 domain, for example, a human IgG4 hinge-Ch2-Ch3 domain.

**[0458]** In some embodiments, the transmembrane domain of the CD20 CAR comprises a CD8α transmembrane domain, for example, a human CD8α transmembrane domain. In some embodiments, the CD8α transmembrane domain comprises or consists of an amino acid sequence set forth in SEQ ID NO:294 or an amino acid sequence that is at least 80% identical (*e.g.,* at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical) to the amino acid sequence set forth in SEQ ID NO:294. In some embodiments, the transmembrane domain comprises a CD28 transmembrane domain, for example, a human CD28 transmembrane domain. In some embodiments, the CD28 transmembrane domain comprises or consists of an amino acid sequence set forth in SEQ ID NO:296 or an amino acid sequence that is at least 80% identical (*e.g.,* at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical) to the amino acid sequence set forth in SEQ ID NO:296.

**[0459]** In some embodiments, the intracellular costimulatory domain of the CD20 CAR comprises a 4-1BB costimulatory domain, for example, a human 4-1BB costimulatory domain. In some embodiments, the 4-1BB costimulatory domain comprises or consists of an amino acid sequence set forth

in SEQ ID NO:298 or an amino acid sequence that is at least 80% identical (*e.g.,* at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical) to the amino acid sequence set forth in SEQ ID NO:298. In some embodiments, the intracellular costimulatory domain comprises a CD28 costimulatory domain, for example, a human CD28 costimulatory domain. In some embodiments, the CD28 costimulatory domain comprises or consists of an amino acid sequence set forth in SEQ ID NO:299 or an amino acid sequence that is at least 80% identical (*e.g.,* at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical) to the amino acid sequence set forth in SEQ ID NO:299.

**[0460]** In some embodiments, the intracellular signaling domain of the CD20 CAR comprises a CD3 zeta (ζ) signaling domain, for example, a human CD3ζ signaling domain. In some embodiments, the CD3ζ signaling domain comprises or consists of an amino acid sequence set forth in SEQ ID NO:300 or an amino acid sequence that is at least 80% identical (*e.g.,* at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical) to the amino acid sequence set forth in SEQ ID NO:300.

**[0461]** In some embodiments, the CAR is a CD20 CAR, including, for example, a CD20 CAR comprising the CD20-specific scFv having sequences set forth in SEQ ID NO: 322, the CD8α hinge domain of SEQ ID NO:289, the CD8α transmembrane domain of SEQ ID NO:294, the 4-1BB costimulatory domain of SEQ ID NO:298, the CD3ζ signaling domain of SEQ ID NO:300, and/or variants (i.e., having a sequence that is at least 80% identical, for example, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99 identical to the disclosed sequence) thereof.

**[0462]** In some embodiments, the CAR is a CD20 CAR, including, for example, a CD20 CAR comprising the CD20-specific scFv having sequences set forth in SEQ ID NO:322, the CD28 hinge domain of SEQ ID NO:290, the CD8α transmembrane domain of SEQ ID NO: 294, the 4-1BB costimulatory domain of SEQ ID NO:298, the CD3ζ signaling domain of SEQ ID NO:300, and/or variants (i.e., having a sequence that is at least 80% identical, for example, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99 identical to the disclosed sequence) thereof.

**[0463]** In some embodiments, the CAR is a CD20 CAR, including, for example, a CD20 CAR comprising the CD20-specific scFv having sequences set forth in SEQ ID NO:322, the IgG4 hinge domain of SEQ ID NO:292 or SEQ ID NO:293, the CD8α transmembrane domain of SEQ ID NO:294, the 4-1BB costimulatory domain of SEQ ID NO: 298, the CD3ζ signaling domain of SEQ ID NO:300, and/or variants (i.e., having a sequence that is at least 80% identical, for example, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99 identical to the disclosed sequence) thereof.

**[0464]** In some embodiments, the CAR is a CD20 CAR, including, for example, a CD20 CAR comprising the CD20-specific scFv having sequences set forth in SEQ ID NO:322, the CD8α hinge domain of SEQ ID NO:289, the CD28 transmembrane domain of SEQ ID NO:296, the 4-1BB costimulatory domain of SEQ ID NO:298, the CD3ζ signaling domain of SEQ ID NO:300, and/or variants (i.e., having a sequence that is at least 80% identical, for example, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99 identical to the disclosed sequence) thereof.

**[0465]** In some embodiments, the CAR is a CD20 CAR, including, for example, a CD20 CAR comprising the CD20-specific scFv having sequences set forth in SEQ ID NO:322, the CD28 hinge domain of SEQ ID NO:290, the CD28 transmembrane domain of SEQ ID NO:296, the 4-1BB costimulatory domain of SEQ ID NO:298, the CD3ζ signaling domain of SEQ ID NO:300, and/or variants (i.e., having a sequence that is at least 80% identical, for example, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99 identical to the disclosed sequence) thereof.

**[0466]** In some embodiments, the CAR is a CD20 CAR, including, for example, a CD20 CAR comprising the CD20-specific scFv having sequences set forth in SEQ ID NO:322, the IgG4 hinge domain of SEQ ID NO:292 or SEQ ID NO:293, the CD28 transmembrane domain of SEQ ID NO:296, the 4-1BB costimulatory domain of SEQ ID NO: 298, the CD3ζ signaling domain of SEQ ID NO:300, and/or variants (i.e., having a sequence that is at least 80% identical, for example, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99 identical to the disclosed sequence) thereof.

### *C) CD22 CAR*

**[0467]** In some embodiments, the CAR is a CD22 CAR ("CD22-CAR"). CD22, which is a transmembrane protein found mostly on the surface of mature B cells that functions as an inhibitory receptor for B cell receptor (BCR) signaling. CD22 is expressed in 60-70% of B cell lymphomas and leukemias (*e.g.,* B-chronic lymphocytic leukemia, hairy cell leukemia, acute lymphocytic leukemia (ALL), and Burkitt's lymphoma) and is not present on the cell surface in early stages of B cell development or on stem cells. In some embodiments, the CD22 CAR may comprise a signal peptide, an extracellular binding domain that specifically binds CD22, a hinge domain, a transmembrane domain, an intracellular costimulatory domain, and/or an intracellular signaling domain in tandem.

**[0468]** In some embodiments, the signal peptide of the CD22 CAR comprises a CD8α signal peptide. In some embodiments, the CD8α signal peptide comprises or consists of an amino acid sequence set forth in SEQ ID NO:286 or an amino acid sequence that is at least 80% identical (*e.g.,* at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical) to the amino acid sequence set forth in of SEQ ID NO:286. In some embodiments,

the signal peptide comprises an IgK signal peptide. In some embodiments, the IgK signal peptide comprises or consists of an amino acid sequence set forth in SEQ ID NO:287 or an amino acid sequence that is at least 80% identical (*e.g.,* at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical) to the amino acid sequence set forth in of SEQ ID NO:287. In some embodiments, the signal peptide comprises a GMCSFR-α or CSF2RA signal peptide. In some embodiments, the GMCSFR-α or CSF2RA signal peptide comprises or consists of an amino acid sequence set forth in SEQ ID NO:288 or an amino acid sequence that is at least 80% identical (*e.g.,* at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical) to the amino acid sequence set forth in of SEQ ID NO:288.

**[0469]** In some embodiments, the extracellular binding domain of the CD22 CAR is specific to CD22, for example, human CD22. The extracellular binding domain of the CD22 CAR can be codon-optimized for expression in a host cell or to have variant sequences to increase functions of the extracellular binding domain. In some embodiments, the extracellular binding domain comprises an immunogenically active portion of an immunoglobulin molecule, for example, an scFv.

**[0470]** In some embodiments, the extracellular binding domain of the CD22 CAR is derived from an antibody specific to CD22, including, for example, SM03, inotuzumab, epratuzumab, moxetumomab, and pinatuzumab. In any of these embodiments, the extracellular binding domain of the CD22 CAR can comprise or consist of the $V_H$, the $V_L$, and/or one or more CDRs of any of the antibodies.

**[0471]** In some embodiments, the extracellular binding domain of the CD22 CAR comprises an scFv derived from the m971 monoclonal antibody (m971), which comprises the heavy chain variable region ($V_H$) and the light chain variable region ($V_L$) of m971 connected by a linker. In some embodiments, the linker is a $3xG_4S$ linker. In other embodiments, the Whitlow linker may be used instead. In some embodiments, the amino acid sequences of the entire m971-derived scFv (also referred to as m971 scFv) and its different portions are provided in **Table 14** below. In some embodiments, the CD22-specific scFv comprises or consists of an amino acid sequence set forth in SEQ ID NO:331, 332, or 336, or an amino acid sequence that is at least 80% identical (*e.g.,* at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical) to the amino acid sequence set forth in of SEQ ID NO: 331, 332, or 336. In some embodiments, the CD22-specific scFv may comprise one or more CDRs having amino acid sequences set forth in SEQ ID NOs: 333-335 and 337-339. In some embodiments, the CD22-specific scFv may comprise a heavy chain with one or more CDRs having amino acid sequences set forth in SEQ ID NOs: 333-335. In some embodiments, the CD22-specific scFv may comprise a light chain with one or more CDRs having amino acid sequences set forth in SEQ ID NOs: 337-339. In any of these embodiments, the CD22-specific scFv may comprise one or more CDRs comprising one or more amino acid substitutions, or comprising a sequence that is at least 80% identical (*e.g.,* at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical), to any of the sequences identified. In some embodiments,

the extracellular binding domain of the CD22 CAR comprises or consists of the one or more CDRs as described herein.

**[0472]** In some embodiments, the extracellular binding domain of the CD22 CAR comprises an scFv derived from m971-L7, which is an affinity matured variant of m971 with significantly improved CD22 binding affinity compared to the parental antibody m971 (improved from about 2 nM to less than 50 pM). In some embodiments, the scFv derived from m971-L7 comprises the $V_H$ and the $V_L$ of m971-L7 connected by a 3x$G_4$S linker. In other embodiments, the Whitlow linker may be used instead. In some embodiments, the amino acid sequences of the entire m971-L7-derived scFv (also referred to as m971-L7 scFv) and its different portions are provided in **Table 14** below. In some embodiments, the CD22-specific scFv comprises or consists of an amino acid sequence set forth in SEQ ID NO:340, 341, or 345, or an amino acid sequence that is at least 80% identical (*e.g.,* at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical) to the amino acid sequence set forth in of SEQ ID NO: 340, 341, or 345. In some embodiments, the CD22-specific scFv may comprise one or more CDRs having amino acid sequences set forth in SEQ ID NOs: 342-344 and 346-348. In some embodiments, the CD22-specific scFv may comprise a heavy chain with one or more CDRs having amino acid sequences set forth in SEQ ID NOs: 342-344. In some embodiments, the CD22-specific scFv may comprise a light chain with one or more CDRs having amino acid sequences set forth in SEQ ID NOs: 346-348. In any of these embodiments, the CD22-specific scFv may comprise one or more CDRs comprising one or more amino acid substitutions, or comprising a sequence that is at least 80% identical (*e.g.,* at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical), to any of the sequences identified. In some embodiments, the extracellular binding domain of the CD22 CAR comprises or consists of the one or more CDRs as described herein.

**Table 14. Exemplary sequences of anti-CD22 scFv and components**

| SEQ ID NO: | Amino Acid Sequence | Description |
|---|---|---|
| 331 | QVQLQQSGPGLVKPSQTLSLTCAISGDS VSSNSAAWNWIRQSPSRGLEWLGRTYY RSKWYNDYAVSVKSRITINPDTSKNQFS LQLNSVTPEDTAVYYCAREVTGDLEDA FDIWGQGTMVTVSSGGGGSGGGGSGG GGSDIQMTQSPSSLSASVGDRVTITCRA SQTIWSYLNWYQQRPGKAPNLLIYAAS SLQSGVPSRFSGRGSGTDFTLTISSLQAE DFATYYCQQSYSIPQTFGQGTKLEIK | Anti-CD22 m971 scFv entire sequence, with 3x$G_4$S linker |
| 332 | QVQLQQSGPGLVKPSQTLSLTCAISGDS VSSNSAAWNWIRQSPSRGLEWLGRTYY RSKWYNDYAVSVKSRITINPDTSKNQFS LQLNSVTPEDTAVYYCAREVTGDLEDA FDIWGQGTMVTVSS | Anti-CD22 m971 scFv heavy chain variable region |

| SEQ ID NO: | Amino Acid Sequence | Description |
|---|---|---|
| 333 | GDSVSSNSAA | Anti-CD22 m971 scFv heavy chain CDR1 |
| 334 | TYYRSKWYN | Anti-CD22 m971 scFv heavy chain CDR2 |
| 335 | AREVTGDLEDAFDI | Anti-CD22 m971 scFv heavy chain CDR3 |
| 336 | DIQMTQSPSSLSASVGDRVTITCRASQTI WSYLNWYQQRPGKAPNLLIYAASSLQS GVPSRFSGRGSGTDFTLTISSLQAEDFAT YYCQQSYSIPQTFGQGTKLEIK | Anti-CD22 m971 scFv light chain |
| 337 | QTIWSY | Anti-CD22 m971 scFv light chain CDR1 |
| 338 | AAS | Anti-CD22 m971 scFv light chain CDR2 |
| 339 | QQSYSIPQT | Anti-CD22 m971 scFv light chain CDR3 |
| 340 | QVQLQQSGPGMVKPSQTLSLTCAISGD SVSSNSVAWNWIRQSPSRGLEWLGRTY YRSTWYNDYAVSMKSRITINPDTNKNQ FSLQLNSVTPEDTAVYYCAREVTGDLE DAFDIWGQGTMVTVSSGGGGSGGGGS GGGGSDIQMIQSPSSLSASVGDRVTITC RASQTIWSYLNWYRQRPGEAPNLLIYA ASSLQSGVPSRFSGRGSGTDFTLTISSLQ AEDFATYYCQQSYSIPQTFGQGTKLEIK | Anti-CD22 m971-L7 scFv entire sequence, with 3x$G_4S$ linker |
| 341 | QVQLQQSGPGMVKPSQTLSLTCAISGD SVSSNSVAWNWIRQSPSRGLEWLGRTY YRSTWYNDYAVSMKSRITINPDTNKNQ FSLQLNSVTPEDTAVYYCAREVTGDLE DAFDIWGQGTMVTVSS | Anti-CD22 m971-L7 scFv heavy chain variable region |
| 342 | GDSVSSNSVA | Anti-CD22 m971-L7 scFv heavy chain CDR1 |
| 343 | TYYRSTWYN | Anti-CD22 m971-L7 scFv heavy chain CDR2 |
| 344 | AREVTGDLEDAFDI | Anti-CD22 m971-L7 scFv heavy chain CDR3 |
| 345 | DIQMIQSPSSLSASVGDRVTITCRASQTI WSYLNWYRQRPGEAPNLLIYAASSLQS GVPSRFSGRGSGTDFTLTISSLQAEDFAT YYCQQSYSIPQTFGQGTKLEIK | Anti-CD22 m971-L7 scFv light chain variable region |
| 346 | QTIWSY | Anti-CD22 m971-L7 scFv light chain CDR1 |
| 347 | AAS | Anti-CD22 m971-L7 scFv light chain CDR2 |

| SEQ ID NO: | Amino Acid Sequence | Description |
|---|---|---|
| 348 | QQSYSIPQT | Anti-CD22 m971-L7 scFv light chain CDR3 |

**[0473]** In some embodiments, the extracellular binding domain of the CD22 CAR comprises immunotoxins HA22 or BL22. Immunotoxins BL22 and HA22 are therapeutic agents that comprise an scFv specific for CD22 fused to a bacterial toxin, and thus can bind to the surface of the cancer cells that express CD22 and kill the cancer cells. BL22 comprises a dsFv of an anti-CD22 antibody, RFB4, fused to a 38-kDa truncated form of *Pseudomonas* exotoxin A (Bang et al., Clin. Cancer Res., 11:1545-50 (2005)). HA22 (CAT8015, moxetumomab pasudotox) is a mutated, higher affinity version of BL22 (Ho et al., J. Biol. Chem., 280(1): 607-17 (2005)). Suitable sequences of antigen binding domains of HA22 and BL22 specific to CD22 are disclosed in, for example, U.S. Patent Nos. 7,541,034; 7,355,012; and 7,982,011, which are hereby incorporated by reference in their entirety.

**[0474]** In some embodiments, the hinge domain of the CD22 CAR comprises a CD8α hinge domain, for example, a human CD8α hinge domain. In some embodiments, the CD8α hinge domain comprises or consists of an amino acid sequence set forth in SEQ ID NO:289 or an amino acid sequence that is at least 80% identical (*e.g.,* at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical) to the amino acid sequence set forth in of SEQ ID NO:289. In some embodiments, the hinge domain comprises a CD28 hinge domain, for example, a human CD28 hinge domain. In some embodiments, the CD28 hinge domain comprises or consists of an amino acid sequence set forth in SEQ ID NO:290 or an amino acid sequence that is at least 80% identical (*e.g.,* at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical) to the amino acid sequence set forth in of SEQ ID NO:290. In some embodiments, the hinge domain comprises an IgG4 hinge domain, for example, a human IgG4 hinge domain. In some embodiments, the IgG4 hinge domain comprises or consists of an amino acid sequence set forth in SEQ ID NO:292 or SEQ ID NO:293, or an amino acid sequence that is at least 80% identical (*e.g.,* at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical) to the amino acid sequence set forth in of SEQ ID NO:292 or SEQ ID NO:293. In some embodiments, the hinge domain comprises a IgG4 hinge-Ch2-Ch3 domain, for example, a human IgG4 hinge-Ch2-Ch3 domain.

**[0475]** In some embodiments, the transmembrane domain of the CD22 CAR comprises a CD8α transmembrane domain, for example, a human CD8α transmembrane domain. In some embodiments, the CD8α transmembrane domain comprises or consists of an amino acid sequence set forth in SEQ ID NO:294 or an amino acid sequence that is at least 80% identical (*e.g.,* at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical) to the amino acid sequence set forth in SEQ ID NO:294. In some embodiments, the transmembrane domain comprises a CD28 transmembrane domain, for example, a human CD28 transmembrane domain. In

some embodiments, the CD28 transmembrane domain comprises or consists of an amino acid sequence set forth in SEQ ID NO:296 or an amino acid sequence that is at least 80% identical (*e.g.,* at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical) to the amino acid sequence set forth in SEQ ID NO:296.

**[0476]** In some embodiments, the intracellular costimulatory domain of the CD22 CAR comprises a 4-1BB costimulatory domain, for example, a human 4-1BB costimulatory domain. In some embodiments, the 4-1BB costimulatory domain comprises or consists of an amino acid sequence set forth in SEQ ID NO:298 or an amino acid sequence that is at least 80% identical (*e.g.,* at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical) to the amino acid sequence set forth in SEQ ID NO:298. In some embodiments, the intracellular costimulatory domain comprises a CD28 costimulatory domain, for example, a human CD28 costimulatory domain. In some embodiments, the CD28 costimulatory domain comprises or consists of an amino acid sequence set forth in SEQ ID NO:299 or an amino acid sequence that is at least 80% identical (*e.g.,* at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical) to the amino acid sequence set forth in SEQ ID NO:299.

**[0477]** In some embodiments, the intracellular signaling domain of the CD22 CAR comprises a CD3 zeta (ζ) signaling domain, for example, a human CD3ζ signaling domain. In some embodiments, the CD3ζ signaling domain comprises or consists of an amino acid sequence set forth in SEQ ID NO:300 or an amino acid sequence that is at least 80% identical (*e.g.,* at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical) to the amino acid sequence set forth in SEQ ID NO:300.

**[0478]** In some embodiments, the CAR is a CD22 CAR, including, for example, a CD22 CAR comprising the CD22-specific scFv having sequences set forth in SEQ ID NO:331 or SEQ ID NO:340, the CD8α hinge domain of SEQ ID NO:289, the CD8α transmembrane domain of SEQ ID NO:295, the 4-1BB costimulatory domain of SEQ ID NO:298, the CD3ζ signaling domain of SEQ ID NO:300, and/or variants (i.e., having a sequence that is at least 80% identical, for example, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99 identical to the disclosed sequence) thereof.

**[0479]** In some embodiments, the CAR is a CD22 CAR, including, for example, a CD22 CAR comprising the CD22-specific scFv having sequences set forth in SEQ ID NO:331 or SEQ ID NO:340, the CD28 hinge domain of SEQ ID NO:290, the CD8α transmembrane domain of SEQ ID NO:295, the 4-1BB costimulatory domain of SEQ ID NO:298, the CD3ζ signaling domain of SEQ ID NO:300, and/or variants (i.e., having a sequence that is at least 80% identical, for example, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99 identical to the disclosed sequence) thereof.

**[0480]** In some embodiments, the CAR is a CD22 CAR, including, for example, a CD22 CAR comprising the CD22-specific scFv having sequences set forth in SEQ ID NO:331 or SEQ ID NO:340, the IgG4 hinge domain of SEQ ID NO:292 or SEQ ID NO:293, the CD8α transmembrane domain of SEQ ID NO:294, the 4-1BB costimulatory domain of SEQ ID NO:298, the CD3ζ signaling domain of SEQ ID NO:300, and/or variants (i.e., having a sequence that is at least 80% identical, for example, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99 identical to the disclosed sequence) thereof.

**[0481]** In some embodiments, the CAR is a CD22 CAR, including, for example, a CD22 CAR comprising the CD22-specific scFv having sequences set forth in SEQ ID NO:331 or SEQ ID NO:340, the CD8α hinge domain of SEQ ID NO:289, the CD28 transmembrane domain of SEQ ID NO:296, the 4-1BB costimulatory domain of SEQ ID NO:298, the CD3ζ signaling domain of SEQ ID NO:300, and/or variants (i.e., having a sequence that is at least 80% identical, for example, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99 identical to the disclosed sequence) thereof.

**[0482]** In some embodiments, the CAR is a CD22 CAR, including, for example, a CD22 CAR comprising the CD22-specific scFv having sequences set forth in SEQ ID NO:331 or SEQ ID NO:340, the CD28 hinge domain of SEQ ID NO:290, the CD28 transmembrane domain of SEQ ID NO:296, the 4-1BB costimulatory domain of SEQ ID NO:298, the CD3ζ signaling domain of SEQ ID NO:300, and/or variants (i.e., having a sequence that is at least 80% identical, for example, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99 identical to the disclosed sequence) thereof.

**[0483]** In some embodiments, the CAR is a CD22 CAR, including, for example, a CD22 CAR comprising the CD22-specific scFv having sequences set forth in SEQ ID NO:331 or SEQ ID NO:340, the IgG4 hinge domain of SEQ ID NO:292 or SEQ ID NO:293, the CD28 transmembrane domain of SEQ ID NO:296, the 4-1BB costimulatory domain of SEQ ID NO:298, the CD3ζ signaling domain of SEQ ID NO:300, and/or variants (i.e., having a sequence that is at least 80% identical, for example, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99 identical to the disclosed sequence) thereof.

### *D) BCMA CAR*

**[0484]** In some embodiments, the CAR is a BCMA CAR ("BCMA-CAR"). BCMA is a tumor necrosis family receptor (TNFR) member expressed on cells of the B cell lineage, with the highest expression on terminally differentiated B cells or mature B lymphocytes. BCMA is involved in mediating the survival of plasma cells for maintaining long-term humoral immunity. The expression of BCMA has been recently linked to a number of cancers, such as multiple myeloma, Hodgkin's and non-Hodgkin's lymphoma, various leukemias, and glioblastoma. In some embodiments, the BCMA CAR

may comprise a signal peptide, an extracellular binding domain that specifically binds BCMA, a hinge domain, a transmembrane domain, an intracellular costimulatory domain, and/or an intracellular signaling domain in tandem.

**[0485]** In some embodiments, the signal peptide of the BCMA CAR comprises a CD8α signal peptide. In some embodiments, the CD8α signal peptide comprises or consists of an amino acid sequence set forth in SEQ ID NO:286 or an amino acid sequence that is at least 80% identical (*e.g.,* at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical) to the amino acid sequence set forth in of SEQ ID NO:286. In some embodiments, the signal peptide comprises an IgK signal peptide. In some embodiments, the IgK signal peptide comprises or consists of an amino acid sequence set forth in SEQ ID NO:287 or an amino acid sequence that is at least 80% identical (*e.g.,* at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical) to the amino acid sequence set forth in of SEQ ID NO:287. In some embodiments, the signal peptide comprises a GMCSFR-α or CSF2RA signal peptide. In some embodiments, the GMCSFR-α or CSF2RA signal peptide comprises or consists of an amino acid sequence set forth in SEQ ID NO:288 or an amino acid sequence that is at least 80% identical (*e.g.,* at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical) to the amino acid sequence set forth in of SEQ ID NO:288.

**[0486]** In some embodiments, the extracellular binding domain of the BCMA CAR is specific to BCMA, for example, human BCMA. The extracellular binding domain of the BCMA CAR can be codon-optimized for expression in a host cell or to have variant sequences to increase functions of the extracellular binding domain.

**[0487]** In some embodiments, the extracellular binding domain comprises an immunogenically active portion of an immunoglobulin molecule, for example, an scFv. In some embodiments, the extracellular binding domain of the BCMA CAR is derived from an antibody specific to BCMA, including, for example, belantamab, erlanatamab, teclistamab, LCAR-B38M, and ciltacabtagene. In any of these embodiments, the extracellular binding domain of the BCMA CAR can comprise or consist of the $V_H$, the $V_L$, and/or one or more CDRs of any of the antibodies.

**[0488]** In some embodiments, the extracellular binding domain of the BCMA CAR comprises an scFv derived from C11D5.3, a murine monoclonal antibody as described in Carpenter et al., Clin. Cancer Res. 19(8):2048-2060 (2013). See also PCT Application Publication No. WO2010/104949. The C11D5.3-derived scFv may comprise the heavy chain variable region ($V_H$) and the light chain variable region ($V_L$) of C11D5.3 connected by the Whitlow linker, the amino acid sequences of which is provided in **Table 15** below. In some embodiments, the BCMA-specific extracellular binding domain comprises or consists of an amino acid sequence set forth in SEQ ID NO:349, 350, or 354, or an amino acid sequence that is at least 80% identical (*e.g.,* at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical) to the amino acid sequence set forth in

of SEQ ID NO:349, 350, or 354. In some embodiments, the BCMA-specific extracellular binding domain may comprise one or more CDRs having amino acid sequences set forth in SEQ ID NOs: 351-353 and 355-357. In some embodiments, the BCMA-specific extracellular binding domain may comprise a light chain with one or more CDRs having amino acid sequences set forth in SEQ ID NOs: 351-353. In some embodiments, the BCMA-specific extracellular binding domain may comprise a heavy chain with one or more CDRs having amino acid sequences set forth in SEQ ID NOs: 355-357. In any of these embodiments, the BCMA-specific scFv may comprise one or more CDRs comprising one or more amino acid substitutions, or comprising a sequence that is at least 80% identical (*e.g.,* at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical), to any of the sequences identified. In some embodiments, the extracellular binding domain of the BCMA CAR comprises or consists of the one or more CDRs as described herein.

**[0489]** In some embodiments, the extracellular binding domain of the BCMA CAR comprises an scFv derived from another murine monoclonal antibody, C12A3.2, as described in Carpenter et al., Clin. Cancer Res. 19(8):2048-2060 (2013) and PCT Application Publication No. WO2010/104949, the amino acid sequence of which is also provided in **Table 15** below. In some embodiments, the BCMA-specific extracellular binding domain comprises or consists of an amino acid sequence set forth in SEQ ID NO:358, 359, 363, or an amino acid sequence that is at least 80% identical (*e.g.,* at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical) to the amino acid sequence set forth in of SEQ ID NO: 358, 359, 363. In some embodiments, the BCMA-specific extracellular binding domain may comprise one or more CDRs having amino acid sequences set forth in SEQ ID NOs: 360-362 and 364-366. In some embodiments, the BCMA-specific extracellular binding domain may comprise a light chain with one or more CDRs having amino acid sequences set forth in SEQ ID NOs: 360-362. In some embodiments, the BCMA-specific extracellular binding domain may comprise a heavy chain with one or more CDRs having amino acid sequences set forth in SEQ ID NOs: 364-366. In any of these embodiments, the BCMA-specific scFv may comprise one or more CDRs comprising one or more amino acid substitutions, or comprising a sequence that is at least 80% identical (*e.g.,* at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical), to any of the sequences identified. In some embodiments, the extracellular binding domain of the BCMA CAR comprises or consists of the one or more CDRs as described herein.

**[0490]** In some embodiments, the extracellular binding domain of the BCMA CAR comprises a murine monoclonal antibody with high specificity to human BCMA, referred to as BB2121 in Friedman et al., Hum. Gene Ther. 29(5):585-601 (2018)). See also, PCT Application Publication No. WO2012163805.

**[0491]** In some embodiments, the extracellular binding domain of the BCMA CAR comprises single variable fragments of two heavy chains (VHH) that can bind to two epitopes of BCMA as described in

Zhao et al., J. Hematol. Oncol. 11(1):141 (2018), also referred to as LCAR-B38M. See also, PCT Application Publication No. WO2018/028647.

**[0492]** In some embodiments, the extracellular binding domain of the BCMA CAR comprises a fully human heavy-chain variable domain (FHVH) as described in Lam et al., Nat. Commun. 11(1):283 (2020), also referred to as FHVH33. See also, PCT Application Publication No. WO2019/006072. The amino acid sequences of FHVH33 and its CDRs are provided in **Table 15** below. In some embodiments, the BCMA-specific extracellular binding domain comprises or consists of an amino acid sequence set forth in SEQ ID NO:367 or an amino acid sequence that is at least 80% identical (*e.g.,* at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical) to the amino acid sequence set forth in of SEQ ID NO:367. In some embodiments, the BCMA-specific extracellular binding domain may comprise one or more CDRs having amino acid sequences set forth in SEQ ID NOs: 368-370. In any of these embodiments, the BCMA-specific extracellular binding domain may comprise one or more CDRs comprising one or more amino acid substitutions, or comprising a sequence that is at least 80% identical (*e.g.,* at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical), to any of the sequences identified. In some embodiments, the extracellular binding domain of the BCMA CAR comprises or consists of the one or more CDRs as described herein.

**[0493]** In some embodiments, the extracellular binding domain of the BCMA CAR comprises an scFv derived from CT103A (or CAR0085) as described in U.S. Patent No. 11,026,975 B2, the amino acid sequence of which is provided in **Table 16** below. In some embodiments, the BCMA-specific extracellular binding domain comprises or consists of an amino acid sequence set forth in SEQ ID NO:371, 372, or 376, or an amino acid sequence that is at least 80% identical (*e.g.,* at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical) to the amino acid sequence set forth in of SEQ ID NO: 371, 372, or 376. In some embodiments, the BCMA-specific extracellular binding domain may comprise one or more CDRs having amino acid sequences set forth in SEQ ID NOs: 373-375 and 377-379. In some embodiments, the BCMA-specific extracellular binding domain may comprise a light chain with one or more CDRs having amino acid sequences set forth in SEQ ID NOs: 373-375. In some embodiments, the BCMA-specific extracellular binding domain may comprise a heavy chain with one or more CDRs having amino acid sequences set forth in SEQ ID NOs: 377-379. In any of these embodiments, the BCMA-specific scFv may comprise one or more CDRs comprising one or more amino acid substitutions, or comprising a sequence that is at least 80% identical (*e.g.,* at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical), to any of the sequences identified. In some embodiments, the extracellular binding domain of the BCMA CAR comprises or consists of the one or more CDRs as described herein.

**[0494]** Additionally, CARs and binders directed to BCMA have been described in U.S. Application Publication Nos. 2020/0246381 A1 and 2020/0339699 A1, the entire contents of each of which are incorporated by reference herein.

**Table 15. Exemplary sequences of anti-BCMA binder and components**

| SEQ ID NO: | Amino Acid Sequence | Description |
|---|---|---|
| 349 | DIVLTQSPASLAMSLGKRATISCRASES VSVIGAHLIHWYQQKPGQPPKLLIYLAS NLETGVPARFSGSGSGTDFTLTIDPVEE DDVAIYSCLQSRIFPRTFGGGTKLEIKGS TSGSGKPGSGEGSTKGQIQLVQSGPELK KPGETVKISCKASGYTFTDYSINWVKR APGKGLKWMGWINTETREPAYAYDFR GRFAFSLETSASTAYLQINNLKYEDTAT YFCALDYSYAMDYWGQGTSVTVSS | Anti-BCMA C11D5.3 scFv entire sequence, with Whitlow linker |
| 350 | DIVLTQSPASLAMSLGKRATISCRASES VSVIGAHLIHWYQQKPGQPPKLLIYLAS NLETGVPARFSGSGSGTDFTLTIDPVEE DDVAIYSCLQSRIFPRTFGGGTKLEIK | Anti-BCMA C11D5.3 scFv light chain variable region |
| 351 | RASESVSVIGAHLIH | Anti-BCMA C11D5.3 scFv light chain CDR1 |
| 352 | LASNLET | Anti-BCMA C11D5.3 scFv light chain CDR2 |
| 353 | LQSRIFPRT | Anti-BCMA C11D5.3 scFv light chain CDR3 |
| 354 | QIQLVQSGPELKKPGETVKISCKASGYT FTDYSINWVKRAPGKGLKWMGWINTE TREPAYAYDFRGRFAFSLETSASTAYLQ INNLKYEDTATYFCALDYSYAMDYWG QGTSVTVSS | Anti-BCMA C11D5.3 scFv heavy chain variable region |
| 355 | DYSIN | Anti-BCMA C11D5.3 scFv heavy chain CDR1 |
| 356 | WINTETREPAYAYDFRG | Anti-BCMA C11D5.3 scFv heavy chain CDR2 |
| 357 | DYSYAMDY | Anti-BCMA C11D5.3 scFv heavy chain CDR3 |
| 358 | DIVLTQSPPSLAMSLGKRATISCRASESV TILGSHLIYWYQQKPGQPPTLLIQLASN VQTGVPARFSGSGSRTDFTLTIDPVEED DVAVYYCLQSRTIPRTFGGGTKLEIKGS TSGSGKPGSGEGSTKGQIQLVQSGPELK KPGETVKISCKASGYTFRHYSMNWVK QAPGKGLKWMGRINTESGVPIYADDFK GRFAFSVETSASTAYLVINNLKDEDTAS YFCSNDYLYSLDFWGQGTALTVSS | Anti-BCMA C12A3.2 scFv entire sequence, with Whitlow linker |

| SEQ ID NO: | Amino Acid Sequence | Description |
|---|---|---|
| 359 | DIVLTQSPPSLAMSLGKRATISCRASESV TILGSHLIYWYQQKPGQPPTLLIQLASN VQTGVPARFSGSGSRTDFTLTIDPVEED DVAVYYCLQSRTIPRTFGGGTKLEIK | Anti-BCMA C12A3.2 scFv light chain variable region |
| 360 | RASESVTILGSHLIY | Anti-BCMA C12A3.2 scFv light chain CDR1 |
| 361 | LASNVQT | Anti-BCMA C12A3.2 scFv light chain CDR2 |
| 362 | LQSRTIPRT | Anti-BCMA C12A3.2 scFv light chain CDR3 |
| 363 | QIQLVQSGPELKKPGETVKISCKASGYT FRHYSMNWVKQAPGKGLKWMGRINTE SGVPIYADDFKGRFAFSVETSASTAYLV INNLKDEDTASYFCSNDYLYSLDFWGQ GTALTVSS | Anti-BCMA C12A3.2 scFv heavy chain variable region |
| 364 | HYSMN | Anti-BCMA C12A3.2 scFv heavy chain CDR1 |
| 365 | RINTESGVPIYADDFKG | Anti-BCMA C12A3.2 scFv heavy chain CDR2 |
| 366 | DYLYSLDF | Anti-BCMA C12A3.2 scFv heavy chain CDR3 |
| 367 | EVQLLESGGGLVQPGGSLRLSCAASGF TFSSYAMSWVRQAPGKGLEWVSSISGS GDYIYYADSVKGRFTISRDISKNTLYLQ MNSLRAEDTAVYYCAKEGTGANSSLA DYRGQGTLVTVSS | Anti-BCMA FHVH33 entire sequence |
| 368 | GFTFSSYA | Anti-BCMA FHVH33 CDR1 |
| 369 | ISGSGDYI | Anti-BCMA FHVH33 CDR2 |
| 370 | AKEGTGANSSLADY | Anti-BCMA FHVH33 CDR3 |
| 371 | DIQMTQSPSSLSASVGDRVTITCRASQSI SSYLNWYQQKPGKAPKLLIYAASSLQS GVPSRFSGSGSGTDFTLTISSLQPEDFAT YYCQQKYDLLTFGGGTKVEIKGSTSGS GKPGSGEGSTKGQLQLQESGPGLVKPS ETLSLTCTVSGGSISSSSYYWGWIRQPP GKGLEWIGSISYSGSTYYNPSLKSRVTIS VDTSKNQFSLKLSSVTAADTAVYYCAR DRGDTILDVWGQGTMVTVSS | Anti-BCMA CT103A scFv entire sequence, with Whitlow linker |
| 372 | DIQMTQSPSSLSASVGDRVTITCRASQSI SSYLNWYQQKPGKAPKLLIYAASSLQS GVPSRFSGSGSGTDFTLTISSLQPEDFAT YYCQQKYDLLTFGGGTKVEIK | Anti-BCMA CT103A scFv light chain variable region |

| SEQ ID NO: | Amino Acid Sequence | Description |
|---|---|---|
| 373 | QSISSY | Anti-BCMA CT103A scFv light chain CDR1 |
| 374 | AAS | Anti-BCMA CT103A scFv light chain CDR2 |
| 375 | QQKYDLLT | Anti-BCMA CT103A scFv light chain CDR3 |
| 376 | QLQLQESGPGLVKPSETLSLTCTVSGGS ISSSSYYWGWIRQPPGKGLEWIGSISYS GSTYYNPSLKSRVTISVDTSKNQFSLKL SSVTAADTAVYYCARDRGDTILDVWG QGTMVTVSS | Anti-BCMA CT103A scFv heavy chain variable region |
| 377 | GGSISSSSYY | Anti-BCMA CT103A scFv heavy chain CDR1 |
| 378 | ISYSGST | Anti-BCMA CT103A scFv heavy chain CDR2 |
| 379 | ARDRGDTILDV | Anti-BCMA CT103A scFv heavy chain CDR3 |

**[0495]** In some embodiments, the hinge domain of the BCMA CAR comprises a CD8α hinge domain, for example, a human CD8α hinge domain. In some embodiments, the CD8α hinge domain comprises or consists of an amino acid sequence set forth in SEQ ID NO:289 or an amino acid sequence that is at least 80% identical (*e.g.,* at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical) to the amino acid sequence set forth in of SEQ ID NO:289. In some embodiments, the hinge domain comprises a CD28 hinge domain, for example, a human CD28 hinge domain. In some embodiments, the CD28 hinge domain comprises or consists of an amino acid sequence set forth in SEQ ID NO:290 or an amino acid sequence that is at least 80% identical (*e.g.,* at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical) to the amino acid sequence set forth in of SEQ ID NO:290. In some embodiments, the hinge domain comprises an IgG4 hinge domain, for example, a human IgG4 hinge domain. In some embodiments, the IgG4 hinge domain comprises or consists of an amino acid sequence set forth in SEQ ID NO:292 or SEQ ID NO:293, or an amino acid sequence that is at least 80% identical (*e.g.,* at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical) to the amino acid sequence set forth in of SEQ ID NO:292 or SEQ ID NO:293. In some embodiments, the hinge domain comprises a IgG4 hinge-Ch2-Ch3 domain, for example, a human IgG4 hinge-Ch2-Ch3 domain.

**[0496]** In some embodiments, the transmembrane domain of the BCMA CAR comprises a CD8α transmembrane domain, for example, a human CD8α transmembrane domain. In some embodiments, the CD8α transmembrane domain comprises or consists of an amino acid sequence set forth in SEQ ID NO:294 or an amino acid sequence that is at least 80% identical (*e.g.,* at least 80%, at least 85%, at least

90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical) to the amino acid sequence set forth in SEQ ID NO:294. In some embodiments, the transmembrane domain comprises a CD28 transmembrane domain, for example, a human CD28 transmembrane domain. In some embodiments, the CD28 transmembrane domain comprises or consists of an amino acid sequence set forth in SEQ ID NO:296 or an amino acid sequence that is at least 80% identical (*e.g.,* at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical) to the amino acid sequence set forth in SEQ ID NO:296.

**[0497]** In some embodiments, the intracellular costimulatory domain of the BCMA CAR comprises a 4-1BB costimulatory domain, for example, a human 4-1BB costimulatory domain. In some embodiments, the 4-1BB costimulatory domain comprises or consists of an amino acid sequence set forth in SEQ ID NO:298 or an amino acid sequence that is at least 80% identical (*e.g.,* at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical) to the amino acid sequence set forth in SEQ ID NO:298. In some embodiments, the intracellular costimulatory domain comprises a CD28 costimulatory domain, for example, a human CD28 costimulatory domain. In some embodiments, the CD28 costimulatory domain comprises or consists of an amino acid sequence set forth in SEQ ID NO:299 or an amino acid sequence that is at least 80% identical (*e.g.,* at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical) to the amino acid sequence set forth in SEQ ID NO:299.

**[0498]** In some embodiments, the intracellular signaling domain of the BCMA CAR comprises a CD3 zeta (ζ) signaling domain, for example, a human CD3ζ signaling domain. In some embodiments, the CD3ζ signaling domain comprises or consists of an amino acid sequence set forth in SEQ ID NO:300 or an amino acid sequence that is at least 80% identical (*e.g.,* at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical) to the amino acid sequence set forth in SEQ ID NO:300.

**[0499]** In some embodiments, the CAR is a BCMA CAR, including, for example, a BCMA CAR comprising any of the BCMA-specific extracellular binding domains as described, the CD8α hinge domain of SEQ ID NO:289, the CD8α transmembrane domain of SEQ ID NO:294, the 4-1BB costimulatory domain of SEQ ID NO:298, the CD3ζ signaling domain of SEQ ID NO:300, and/or variants (i.e., having a sequence that is at least 80% identical, for example, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99 identical to the disclosed sequence) thereof. In any of these embodiments, the BCMA CAR may additionally comprise a signal peptide (*e.g.,* a CD8α signal peptide) as described.

**[0500]** In some embodiments, the CAR is a BCMA CAR, including, for example, a BCMA CAR comprising any of the BCMA-specific extracellular binding domains as described, the CD8α hinge domain of SEQ ID NO:289, the CD8α transmembrane domain of SEQ ID NO:294, the CD28 costimulatory domain of SEQ ID NO:299, the CD3ζ signaling domain of SEQ ID NO:300, and/or

variants (i.e., having a sequence that is at least 80% identical, for example, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99 identical to the disclosed sequence) thereof. In any of these embodiments, the BCMA CAR may additionally comprise a signal peptide as described.

**[0501]** In some embodiments, the CAR is a BCMA CAR as set forth in SEQ ID NO:380 or is at least 80% identical (*e.g.,* at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical) to the nucleotide sequence set forth in SEQ ID NO:380 (see **Table 16**). The encoded BCMA CAR has a corresponding amino acid sequence set forth in SEQ ID NO:381 or is at least 80% identical (*e.g.,* at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical) to the amino acid sequence set forth in of SEQ ID NO:381, with the following components: CD8α signal peptide, CT103A scFv ($V_L$-Whitlow linker-$V_H$), CD8α hinge domain, CD8α transmembrane domain, 4-1BB costimulatory domain, and CD3ζ signaling domain.

**[0502]** In some embodiments, the CAR is a commercially available embodiment of BCMA CAR, including, for example, idecabtagene vicleucel (ide-cel, also called bb2121). In some embodiments, the CAR is idecabtagene vicleucel or portions thereof. Idecabtagene vicleucel comprises a BCMA CAR with the following components: the BB2121 binder, CD8α hinge domain, CD8α transmembrane domain, 4-1BB costimulatory domain, and CD3ζ signaling domain.

**Table 16. Exemplary sequences of BCMA CARs**

| SEQ ID NO: | Sequence | Description |
|---|---|---|
| 380 | atggccttaccagtgaccgccttgctcctgccgctggccttgctgctcca<br>cgccgccaggccggacatccagatgacccagtctccatcctccctgtct<br>gcatctgtaggagacagagtcaccatcacttgccgggcaagtcagagc<br>attagcagctatttaaattggtatcagcagaaaccagggaaagcccctaa<br>gctcctgatctatgctgcatccagtttgcaaagtggggtcccatcaaggtt<br>cagtggcagtggatctgggacagatttcactctcaccatcagcagtctgc<br>aacctgaagattttgcaacttactactgtcagcaaaaatacgacctcctca<br>cttttggcggagggaccaaggttgagatcaaaggcagcaccagcggct<br>ccggcaagcctggctctggcgagggcagcacaaagggacagctgca<br>gctgcaggagtcgggcccaggactggtgaagccttcggagaccctgtc<br>cctcacctgcactgtctctggtggctccatcagcagtagtagttactactg<br>gggctggatccgccagccccccagggaaggggctggagtggattggg<br>agtatctcctatagtgggagcacctactacaacccgtccctcaagagtcg<br>agtcaccatatccgtagacacgtccaagaaccagttctccctgaagctga<br>gttctgtgaccgccgcagacacggcggtgtactactgcgccagagatc<br>gtggagacaccatactagacgtatggggtcagggtacaatggtcaccgt<br>cagctcattcgtgcccgtgttcctgcccgccaaacctaccaccacccctg<br>cccctagacctcccaccccagccccaacaatcgccagccagcctctgt<br>ctctgcggcccgaagcctgtagacctgctgccggcggagccgtgcaca<br>ccagaggcctggacttcgcctgcgacatctacatctgggcccctctggc<br>cggcacctgtggcgtgctgctgctgagcctggtgatcaccctgtactgc<br>aaccaccggaacaaacggggcagaaagaaactcctgtatatattcaaa<br>caaccatttatgagaccagtacaaactactcaagaggaagatggctgta | Exemplary BCMA CAR nucleotide sequence |

| SEQ ID NO: | Sequence | Description |
|---|---|---|
| | gctgccgatttccagaagaagaagaaggaggatgtgaactgagagtga agttcagcagatccgccgacgcccctgcctaccagcagggacagaac cagctgtacaacgagctgaacctgggcagacgggaagagtacgacgt gctggacaagcggagaggccgggaccccgagatgggcggaaagcc cagacggaagaacccccaggaaggcctgtataacgaactgcagaaag acaagatggccgaggcctacagcgagatcggcatgaagggcgagcg gaggcgcggcaagggccacgatggcctgtaccagggcctgagcacc gccaccaaggacacctacgacgccctgcacatgcaggccctgccccc caga | |
| 381 | MALPVTALLLPLALLLHAARPDIQMTQSPSSL SASVGDRVTITCRASQSISSYLNWYQQKPGKA PKLLIYAASSLQSGVPSRFSGSGSGTDFTLTISS LQPEDFATYYCQQKYDLLTFGGGTKVEIKGST SGSGKPGSGEGSTKGQLQLQESGPGLVKPSET LSLTCTVSGGSISSSSYYWGWIRQPPGKGLEWI GSISYSGSTYYNPSLKSRVTISVDTSKNQFSLK LSSVTAADTAVYYCARDRGDTILDVWGQGT MVTVSSFVPVFLPAKPTTTPAPRPPTPAPTIAS QPLSLRPEACRPAAGGAVHTRGLDFACDIYIW APLAGTCGVLLLSLVITLYCNHRNKRGRKKLL YIFKQPFMRPVQTTQEEDGCSCRFPEEEEGGC ELRVKFSRSADAPAYQQGQNQLYNELNLGRR EEYDVLDKRRGRDPEMGGKPRRKNPQEGLY NELQKDKMAEAYSEIGMKGERRRGKGHDGL YQGLSTATKDTYDALHMQALPPR | Exemplary BCMA CAR amino acid sequence |

### *3. Gene Editing Enzymes*

**[0503]** In some embodiments, the exogenous agent is or comprises a genome editing technology. In some embodiments, the exogenous agent is or comprises a heterologous protein that is associated with a genome editing technology. Any of a variety of agents associated with gene editing technologies can be included as the exogenous agent and/or heterologous protein, such as for delivery of gene editing machinery to a cell. In some embodiments, the gene editing technology can include systems involving nuclease, nickase, homing, integrase, transposase, recombinase, and/or reverse transcriptase activity. In some embodiments, the gene editing technologies can be used for knock-out or knock-down of genes. In some embodiments, the gene-editing technologies can be used for knock-in or integration of DNA into a region of the genome. In some embodiments, the exogenous agent and/or heterologous protein mediates single-strand breaks (SSB). In some embodiments, the exogenous agent and/or heterologous protein mediates double-strand breaks (DSB), including in connection with non-homologous end-joining (NHEJ) or homology-directed repair (HDR). In some embodiments, the exogenous agent and/or heterologous protein does not mediate SSB. In some embodiments, the exogenous agent and/or heterologous protein does not mediate DSB. In some embodiments, the exogenous agent and/or heterologous protein can be used for DNA base editing or prime-editing. In some embodiments, the exogenous agent and/or

heterologous protein can be used for Programmable Addition via Site-specific Targeting Elements (PASTE).

**[0504]** In some embodiments, the exogenous agent is a nuclease for use in gene editing methods. In some embodiments, the nuclease is a zinc-finger nucleases (ZFNs), transcription-activator like effector nucleases (TALENs), or a CRISPR-associated protein- nuclease (Cas). In some embodiments, the Cas protein is selected from the group consisting of Cas3, Cas9, Cas10, Cas12, and Cas13. In some embodiments, the Cas is a Cas12a (also known as cpf1) from a Prevotella, Francisella novicida, Acidaminococcus sp., Lachnospiraceae bacterium, or Francisella bacteria. In some embodiments, the Cas is Cas9 from *Streptococcus pyogenes*. In some embodiments, the Cas is Cas9 from Streptococcus pyogenes (SpCas). In some embodiments, the Cas9 is from Staphylococcus aureus (SaCas9). In some embodiments, the Cas9 is from Neisseria meningitidis (NmeCas9). In some embodiments, the Cas9 is from Campylobacter jejuni (CjCas9). In some embodiments, the Cas9 is from Streptococcus thermophilis (StCas9). In some embodiments, the Cas is a Cas12a (also known as Cpf1) from a *Prevotella* or *Francisella* bacteria, or the Cas is a Cas12b from a Bacillus, optionally *Bacillus hisashii*. In some embodiments, the Cas is a Cas12a (also known as cpf1) from a Prevotella, Francisella novicida, Acidaminococcus sp., Lachnospiraceae bacterium, or Francisella bacteria. In some embodiments, the nuclease is MAD7 or CasX. In some of any embodiments, the Cas is a Cas3, Cas13, CasMini, or any other Cas protein known in the art. See for example, Wang et al., Biosensors and Bioelectronics (165) 1: 2020, and Wu et al. Nature Reviews Chemistry (4) 441: 2020). The Cas9 nuclease can, in some embodiments, be a Cas9 or functional fragment thereof from any bacterial species. See, e.g., Makarova et al. Nature Reviews, Microbiology, 9: 467-477 (2011), including supplemental information, hereby incorporated by reference in its entirety.

**[0505]** In some embodiments, delivery of the nuclease is by a provided vector encoding the nuclease (e.g. Cas).

**[0506]** In some embodiments, the provided viral vector particles contain a nuclease protein and the nuclease protein is directly delivered to a target cell. Methods of delivering a nuclease protein include those as described, for example, in Cai et al. Elife, 2014, 3:e01911 and International patent publication No. WO2017068077. For instance, provided viral vector particles comprise one or more Cas protein(s), such as Cas9. In some embodiments, the nuclease protein (e.g. Cas, such as Cas 9) is engineered as a chimeric nuclease protein with a viral structural protein (e.g. GAG) for packaging into the viral vector particle (e.g. lentiviral vector particle). For instance, a chimeric Cas9-protein fusion with the structural GAG protein can be packaged inside a lentiviral vector particle. In some embodiments, the fusion protein is a cleavable fusion protein between (i) a viral structural protein (e.g. GAG) and (ii) a nuclease protein (e.g. Cas protein, such as Cas 9).

**[0507]** In some embodiments, the Cas is wild-type Cas9, which can site-specifically cleave double-stranded DNA, resulting in the activation of the double-strand break (DSB) repair machinery. DSBs can

be repaired by the cellular Non-Homologous End Joining (NHEJ) pathway (Overballe-Petersen et al., 2013, Proc Natl Acad Sci USA, Vol. 110: 19860-19865), resulting in insertions and/or deletions (indels) which disrupt the targeted locus. Alternatively, if a donor template with homology to the targeted locus is supplied, the DSB may be repaired by the homology-directed repair (HDR) pathway allowing for precise replacement mutations to be made (Overballe- Petersen et al., 2013, Proc Natl Acad Sci USA, Vol. 110: 19860-19865; Gong et al., 2005, Nat. Struct Mol Biol, Vol. 12: 304-312). In some embodiments, the Cas is mutant form, known as Cas9 D10A, with only nickase activity. This means that Cas9D10A cleaves only one DNA strand, and does not activate NHEJ. Instead, when provided with a homologous repair template, DNA repairs are conducted via the high-fidelity HDR pathway only, resulting in reduced indel mutations (Cong et al., 2013, Science, Vol. 339: 819-823; Jinek et al., 2012, Science, Vol.337: 816-821; Qi et al., 2013 Cell, Vol. 152: 1173-1183). Cas9D10A is even more appealing in terms of target specificity when loci are targeted by paired Cas9 complexes designed to generate adjacent DNA nicks (Ran et al., 2013, Cell, Vol. 154: 1380-1389). In some embodiments, the Cas is a nuclease-deficient Cas9 (Qi et al., 2013 Cell, Vol. 152: 1173-1183). For instance, mutations H840A in the HNH domain and D10A in the RuvC domain inactivate cleavage activity, but do not prevent DNA binding. Therefore, this variant can be used to target in a sequence-specific manner any region of the genome without cleavage. Instead, by fusing with various effector domains, dCas9 can be used either as a gene silencing or activation tools. Furthermore, it can be used as a visualization tool by coupling the guide RNA or the Cas9 protein to a fluorophore or a fluorescent protein. In some embodiments, the Cas protein comprises one or more mutations such that the Cas protein is converted into a nickase that is able to cleave only one strand of a double stranded DNA molecule (e.g., a SSB). In some embodiments, the Cas protein is selected from the group consisting of Cas3, Cas4, Cas5, Cas8a, Cas8b, Cas8c, Cas9, Cas10, Cas12, Cas12a (Cpf1), Cas12b (C2c1), Cas12c (C2c3), Cas12d (CasY), Cas12e (CasX), Cas12f (C2c10), Cas12g, Cas12h, Cas12i, Cas12k (C2c5), Cas13, Cas13a (C2c2), Cas13b, Cas13c, Cas13d, C2c4, C2c8, C2c9, Cmr5, Cse1, Cse2, Csf1, Csm2, Csn2, Csx10, Csx11, Csy1, Csy2, Csy3, and Mad7. In some embodiments, the Cas protein is Cas9. In some embodiments, the Cas9 is from a bacteria selected from the group consisting of Streptococcus pyogenes, Staphylococcus aureus, Neisseria meningitides, Campylobacter jejuni, and Streptococcus thermophilis. In some embodiments, the Cas9 is from Streptococcus pyogenes. In some embodiments, the Cas9 is from Streptococcus pyogenes and comprises one or more mutations in the RuvC I, RuvC II, or RuvC III motifs. In some embodiments, the Cas9 is from Streptococcus pyogenes and comprises a D10A mutation in the RuvC I motif. In some embodiments, the Cas9 is from Streptococcus pyogenes and comprises one or more mutations in the HNH catalytic domain. In some embodiments, the Cas9 is from Streptococcus pyogenes and comprises one or more mutations in the HNH catalytic domain selected from the group consisting of H840A, H854A, and H863A. In some embodiments, the Cas9 is from Streptococcus pyogenes and comprises a H840A mutation in the HNH catalytic domain. In some embodiments, the Cas9 is from Streptococcus

pyogenes and comprises a mutation selected from the group consisting of D10A, H840A, H854A, and H863A.

**[0508]** In some embodiments, the Cas protein is selected from the group consisting of Cas3, Cas9, Cas10, Cas12, and Cas13. In particular embodiments, the nuclease is a Cas nuclease, such as Cas9. In some embodiments, delivery of the CRISPR/Cas can be used to introduce single point mutations (deletions or insertions) in a particular target gene, via a single gRNA. Using a pair of gRNA-directed Cas9 nucleases instead, it is also possible to induce large deletions or genomic rearrangements, such as inversions or translocations. In some embodiments, the one or more agent(s) (e.g., the heterologous protein) capable of inducing a DSB comprise Cas9 or a functional fragment thereof, and a first guide RNA, e.g., a first sgRNA, and a second guide RNA, e.g., a second sgRNA. The guide RNA, e.g., the first guide RNA or the second guide RNA, in some embodiments, binds to the recombinant nuclease and targets the recombinant nuclease to a specific location within the target gene such as at a location within the sense strand or the antisense strand of the target gene that is or includes the cleavage site. In some embodiments, the recombinant nuclease is a Cas protein from any bacterial species, or is a functional fragment thereof. In some embodiments, the Cas protein is Cas9 nuclease. Cas9 can, in some embodiments, be a Cas9 or functional fragment thereof from any bacterial species. See, e.g., Makarova et al. Nature Reviews, Microbiology, 9: 467-477 (2011), including supplemental information, hereby incorporated by reference in its entirety. In some embodiments, the Cas9 is from Streptococcus pyogenes (SpCas9). In some embodiments, the Cas9 is from Staphylococcus aureus (SaCas9). In some embodiments, the Cas9 is from Neisseria meningitidis (NmeCas9). In some embodiments, the Cas9 is from Campylobacter jejuni (CjCas9). In some embodiments, the Cas9 is from Streptococcus thermophilis (StCas9).

**[0509]** In some embodiments, the Cas9 is from Streptococcus pyogenes (SpCas9) and comprises one or more mutations in the RuvC catalytic domain or the HNH catalytic domain. In some embodiments, the one or more mutations in the RuvC catalytic domain or the HNH catalytic domain inactivates the catalytic activity of the domain. In some embodiments, the recombinant nuclease has RuvC activity but does not have HNH activity. In some embodiments, the recombinant nuclease does not have RuvC activity but does have HNH activity. In some embodiments, the Cas9 is from Streptococcus pyogenes (SpCas9) and comprises one or more mutations selected from the group consisting of D10A, H840A, H854A, and H863A. In some embodiments, the Cas9 is from Streptococcus pyogenes (SpCas9) and comprises one or more mutations in the RuvC I, RuvC II, or RuvC III motifs. In some embodiments, the Cas9 is from Streptococcus pyogenes (SpCas9) and comprises a mutation in the RuvC I motif. In some embodiments, the Cas9 is from Streptococcus pyogenes (SpCas9) and comprises a D10A mutation in the RuvC I motif. In some embodiments, the Cas9 is from Streptococcus pyogenes (SpCas9) and comprises one or more mutations in the HNH catalytic domain. In some embodiments, the one or more mutations in the HNH catalytic domain is selected from the group

consisting of H840A, H854A, and H863A. In some embodiments, the Cas9 is from Streptococcus pyogenes (SpCas9) and comprises a H840A mutation in the HNH catalytic domain. In some embodiments, the Cas9 is from Streptococcus pyogenes (SpCas9) and comprises a H840A mutation. In some embodiments, the Cas9 is from Streptococcus pyogenes (SpCas9) and comprises a D10A mutation. In some embodiments, the Cas9 is from Streptococcus pyogenes (SpCas9) and comprises one or more mutations selected from the group consisting of N497A, R661A, Q695A, and Q926A. In some embodiments, the Cas9 is from Streptococcus pyogenes (SpCas9) and comprises one or more mutations selected from the group consisting of R780A, K810A, K855A, H982A, K1003A, R1060A, and K848A. In some embodiments, the Cas9 is from Streptococcus pyogenes (SpCas9) and comprises one or more mutations selected from the group consisting of N692A, M694A, Q695A, and H698A. In some embodiments, the Cas9 is from Streptococcus pyogenes (SpCas9) and comprises one or more mutations selected from the group consisting of M495V, Y515N, K526E, and R661Q. In some embodiments, the Cas9 is from Streptococcus pyogenes (SpCas9) and comprises one or more mutations selected from the group consisting of F539S, M763I, and K890N. In some embodiments, the Cas9 is from Streptococcus pyogenes (SpCas9) and comprises one or more mutations selected from the group consisting of E480K, E543D, E1219V, A262T, S409I, M694I, E108G, S217A.

**[0510]** In some embodiments, the Cas9 is from Streptococcus pyogenes (SaCas9). In some embodiments, the SaCas9 is wild type SaCas9. In some embodiments, the SaCas9 comprises one or more mutations in REC3 domain. In some embodiments, the SaCas9 comprises one or more mutations in REC1 domain. In some embodiments, the SaCas9 comprises one or more mutations selected from the group consisting of N260D, N260Q, N260E, Q414A, Q414L. In some embodiments, the SaCas9 comprises one or more mutations in the recognition lobe. In some embodiments, the SaCas9 comprises one or more mutations selected from the group consisting of R245A, N413A, N419A. In some embodiments, the SaCas9 comprises one or more mutations in the RuvC-III domain. In some embodiments, the SaCas9 comprises a R654A mutation.

**[0511]** In some embodiments, the Cas protein is Cas12. In some embodiments, the Cas protein is Cas12a (i.e. cpf1). In some embodiments, the Cas12a is from the group consisting of Francisella novicida U112 (FnCas12a), Acidaminococcus sp. BV3L6 (AsCas12a), Moraxella bovoculi AAX11_00205 (Mb3Cas12a), Lachnospiraceae bacterium ND2006 (LbCas12a), Thiomicrospira sp. Xs5 (TsCas12a), Moraxella bovoculi AAX08_00205 (Mb2Cas12a), and Butyrivibrio sp. NC3005 (BsCas12a). In some embodiments, the Cas12a recognizes a T-rich 5' protospacer adjacent motif (PAM). In some embodiments, the Cas12a processes its own crRNA without requiring a transactivating crRNA (tracrRNA). In some embodiments, the Cas12a processes both RNase and DNase activity. In some embodiments, the Cas12a is a split Cas12a platform, consisting of N-terminal and C-terminal fragments of Cas12a. In some embodiments, the split Cas12a platform is from Lachnospiraceae bacterium.

**[0512]** In some embodiments, the lipid particle further comprises a polynucleotide per se, i.e. a polynucleotide that does not encode for a heterologous protein. In some embodiments, the polynucleotide per se is associated with a gene editing system. For example, a lipid particle may comprise a guide RNA (gRNA), such as a single guide RNA (sgRNA).

**[0513]** In some embodiments, the one or more agent(s) (e.g., one or more exogenous agent and/or heterologous protein) comprise, or are used in combination with, a guide RNA, e.g., single guide RNA (sgRNA), for inducing a DSB at the cleavage site. In some embodiments, the one or more agent(s) comprise, or are used in combination with, more than one guide RNA, e.g., a first sgRNA and a second sgRNA, for inducing a DSB at the cleavage site through a SSB on each strand. In some embodiments, the one or more agent(s) (e.g., the heterologous protein) can be used in combination with a donor template, e.g., a single-stranded DNA oligonucleotide (ssODN), for HDR-mediated integration of the donor template into the target gene, such as at the targeting sequence. In some embodiments, the one or more agent(s) (e.g., one or more exogenous agent and/or heterologous protein) can be used in combination with a donor template, e.g., an ssODN, and a guide RNA, e.g., a sgRNA, for HDR-mediated integration of the donor template into the target gene, such as at the targeting sequence. In some embodiments, the one or more agent(s) (e.g., one or more exogenous agent and/or heterologous protein) can be used in combination with a donor template, e.g., an ssODN, and a first guide RNA, e.g., a first sgRNA, and a second guide RNA, e.g., a second sgRNA, for HDR-mediated integration of the donor template into the target gene, such as at the targeting sequence.

**[0514]** In particular embodiments, the genome-modifying agent is a Cas protein, such as Cas9. In some embodiments, delivery of the CRISPR/Cas can be used to introduce single point mutations (deletions or insertions) in a particular target gene, via a single gRNA. Using a pair of gRNA-directed Cas9 nucleases instead, it is also possible to induce large deletions or genomic rearrangements, such as inversions or translocations. In some embodiments, a dCas9 version of the CRISPR/Cas9 system can be used to target protein domains for transcriptional regulation, epigenetic modification, and microscopic visualization of specific genome loci.

**[0515]** In some embodiments, the genome-modifying agent, e.g., Cas9, is targeted to the cleavage site by interacting with a guide RNA, e.g., sgRNA, that hybridizes to a DNA sequence that immediately precedes a Protospacer Adjacent Motif (PAM) sequence. In general, a guide RNA, e.g., sgRNA, is any nucleotide sequence comprising a sequence, e.g., a crRNA sequence, that has sufficient complementarity with a target gene sequence to hybridize with the target gene sequence at the cleavage site and direct sequence-specific binding of the recombinant nuclease to a portion of the target gene that includes the cleavage site. Full complementarity (100%) is not necessarily required, so long as there is sufficient complementarity to cause hybridization and promote formation of a complex, e.g., CRISPR complex, that includes the recombinant nuclease, e.g., Cas9, and the guide RNA, e.g., sgRNA. In some embodiments, the cleavage site is situated at a site within the target gene that is homologous to the

sequence of the guide RNA, e.g., sgRNA. In some embodiments, the cleavage site is situated approximately 3 nucleotides upstream of the PAM sequence. In some embodiments, the cleavage site is situated approximately 3 nucleotides upstream of the juncture between the guide RNA and the PAM sequence. In some embodiments, the cleavage site is situated 3 nucleotides upstream of the PAM sequence. In some embodiments, the cleavage site is situated 4 nucleotides upstream of the PAM sequence.

**[0516]** In some embodiments, the one or more agent(s) (e.g., one or more exogenous agent and/or heterologous protein) capable of inducing a DSB comprise a fusion protein comprising a DNA binding domain and a DNA cleavage domain. In some embodiments, the DNA cleavage domain is or comprises a recombinant nuclease. In some embodiments, the fusion protein is a TALEN comprising a DNA binding domain and a DNA cleavage domain. In some embodiments, the DNA binding domain is a transcription activator-like (TAL) effector DNA binding domain. In some embodiments, the TAL effector DNA binding domain is from Xanthomonas bacteria. In some embodiments, the DNA cleavage domain is a FokI nuclease domain. In some embodiments, the TAL effector DNA binding domain is engineered to target a specific target sequence, e.g., a portion of a target gene that includes a cleavage site.

**[0517]** In some embodiments, the fusion protein is a zinc finger nuclease (ZFN) comprising a zinc finger DNA binding domain and a DNA cleavage domain. In some embodiments, the DNA cleavage domain is a FokI nuclease domain. In some embodiments, the zinc finger DNA binding domain is engineered to target a specific target sequence, e.g., a portion of a target gene, that includes a cleavage site, such as the targeting sequence.

**[0518]** In some embodiments, the provided lipid particles can be for use in a method to deliver an exogenous agent which involves introducing, into a cell, one or more agent(s) (e.g., one or more exogenous agent and/or heterologous protein) capable of inducing a SSB at a cleavage site within the sense strand and a SSB at a cleavage site within the antisense strand of an endogenous target gene in the cell.

**[0519]** In some embodiments, the cleavage site in the sense strand is less than 400, less than 350, less than 300, less than 250, less than 200, less than 175, less than 150, less than 125, less than 100, less than 90, less than 80, less than 75, less than 70, less than 65, less than 60, less than 55, less than 50, less than 45, less than 40, or less than 35 nucleotides from the nucleotide that is complementary to the cleavage site in the antisense strand. In some embodiments, the cleavage site in the antisense strand is less than 400, less than 350, less than 300, less than 250, less than 200, less than 175, less than 150, less than 125, less than 100, less than 90, less than 80, less than 75, less than 70, less than 65, less than 60, less than 55, less than 50, less than 45, less than 40, or less than 35 nucleotides from the nucleotide that is complementary to the cleavage site in the sense strand. In some embodiments, the cleavage site in the sense strand is between 20 and 400, 20 and 350, 20 and 300, 20 and 250, 20 and 200, 20 and 150, 20 and

125, 20 and 100, 20 and 90, 20 and 80, 20 and 70, 30 and 400, 30 and 350, 30 and 300, 30 and 250, 30 and 200, 30 and 150, 30 and 125, 30 and 100, 30 and 90, 30 and 80, 30 and 70, 40 and 400, 40 and 350, 40 and 300, 40 and 250, 40 and 200, 40 and 150, 40 and 125, 40 and 100, 40 and 90, 40 and 80, or 40 and 70 nucleotides from the nucleotide that is complementary to the cleavage site in the antisense strand. In some embodiments, the cleavage site in the antisense strand is between 20 and 400, 20 and 350, 20 and 300, 20 and 250, 20 and 200, 20 and 150, 20 and 125, 20 and 100, 20 and 90, 20 and 80, 20 and 70, 30 and 400, 30 and 350, 30 and 300, 30 and 250, 30 and 200, 30 and 150, 30 and 125, 30 and 100, 30 and 90, 30 and 80, 30 and 70, 40 and 400, 40 and 350, 40 and 300, 40 and 250, 40 and 200, 40 and 150, 40 and 125, 40 and 100, 40 and 90, 40 and 80, or 40 and 70 nucleotides from the nucleotide that is complementary to the cleavage site in the sense strand.

**[0520]** In some embodiments, the one or more agent(s) (e.g., one or more exogenous agent and/or heterologous protein) capable of inducing a SSB at a cleavage site within the sense strand and a SSB at a cleavage site within the antisense strand comprise a recombinant nuclease. In some embodiments, the recombinant nuclease includes a recombinant nuclease that induces the SSB in the sense strand, and a recombinant nuclease that induced the SSB in the antisense strand, and both of which recombinant nucleases are referred to as the recombinant nuclease. Accordingly, in some embodiments, the method involves introducing, into a cell, one or more agent(s) (e.g., the one or more exogenous agent and/or heterologous protein) comprising a recombinant nuclease for inducing a SSB at a cleavage site in the sense strand and a SSB at a cleavage site in the antisense strand within an endogenous target gene in the cell. Although, in some embodiments, it is described that "a" or "the" recombinant nuclease induces a SSB in the antisense strand a SSB in the sense strand, it is to be understood that this includes situations where two of the same recombinant nuclease is used, such that one of the recombinant nuclease induces the SSB in the sense strand and the other recombinant nuclease induces the SSB in the antisense strand. In some embodiments, the recombinant nuclease that induces the SSB lacks the ability to induce a DSB by cleaving both strands of double stranded DNA.

**[0521]** In some embodiments, the one or more agent(s) capable of inducing a SSB comprise a recombinant nuclease and a first guide RNA, e.g., a first sgRNA, and a second guide RNA, e.g., a second sgRNA.

**[0522]** In some embodiments, the genome-modifying agent is a Cas protein, a transcription activator-like effector nuclease (TALEN), or a zinc finger nuclease (ZFN). In some embodiments, the recombinant nuclease is a Cas nuclease. In some embodiments, the recombinant nuclease is a TALEN. In some embodiments, the recombinant nuclease is a ZFN.

**[0523]** In some embodiments, the one or more agent(s) capable of inducing a SSB at a cleavage site within the sense strand and a SSB at a cleavage site within the antisense strand comprise a fusion protein comprising a DNA binding domain and a DNA cleavage domain. In some embodiments, the DNA cleavage domain is or comprises a recombinant nuclease. In some embodiments, the fusion protein

is a TALEN comprising a DNA binding domain and a DNA cleavage domain. In some embodiments, the DNA binding domain is a transcription activator-like (TAL) effector DNA binding domain. In some embodiments, the TAL effector DNA binding domain is from Xanthomonas bacteria. In some embodiments, the DNA cleavage domain is a FokI nuclease domain. In some embodiments, the TAL effector DNA binding domain is engineered to target a specific target sequence, e.g., a portion of a target gene that includes a cleavage site. In some embodiments, the fusion protein is a zinc finger nuclease (ZFN) comprising a zinc finger DNA binding domain and a DNA cleavage domain. In some embodiments, the DNA cleavage domain is a FokI nuclease domain. In some embodiments, the zinc finger DNA binding domain is engineered to target a specific target sequence, e.g., a portion of a target gene that includes a cleavage site, such as the targeting sequence.

**[0524]** In some embodiments, the one or more agent(s) capable of inducing a SSB at a cleavage site within the sense strand and a SSB at a cleavage site within the antisense strand involve use of the CRISPR/Cas gene editing system. In some embodiments, the one or more agent(s) comprise a recombinant nuclease.

**[0525]** In some embodiments, the genome-modifying agent is a Cas protein. In some embodiments, the Cas protein comprises one or more mutations such that the Cas protein is converted into a nickase that lacks the ability to cleave both strands of a double stranded DNA molecule. In some embodiments, the Cas protein comprises one or more mutations such that the Cas protein is converted into a nickase that is able to cleave only one strand of a double stranded DNA molecule. For example, Cas9, which is normally capable of inducing a double strand break, can be converted into a Cas9 nickase, which is capable of inducing a single strand break, by mutating one of two Cas9 catalytic domains: the RuvC domain, which comprises the RuvC I, RuvC II, and RuvC III motifs, or the NHN domain. In some embodiments, the Cas protein comprises one or more mutations in the RuvC catalytic domain or the HNH catalytic domain. In some embodiments, the genome-modifying protein is a recombinant nuclease that has been modified to have nickase activity. In some embodiments, the recombinant nuclease cleaves the strand to which the guide RNA, e.g., sgRNA, hybridizes, but does not cleave the strand that is complementary to the strand to which the guide RNA, e.g., sgRNA, hybridizes. In some embodiments, the recombinant nuclease does not cleave the strand to which the guide RNA, e.g., sgRNA, hybridizes, but does cleave the strand that is complementary to the strand to which the guide RNA, e.g., sgRNA, hybridizes.

**[0526]** In some embodiments, the lipid particle further comprises a guide RNA (gRNA), such as a single guide RNA (sgRNA). Thus, in some embodiments, the heterologous agent comprises a guide RNA (gRNA). In some embodiments, the gRNA is a single guide RNA (sgRNA).

**[0527]** In some embodiments, the genome-modifying protein, e.g., Cas9, is targeted to the cleavage site by interacting with a guide RNA, e.g., a first guide RNA, such as a first sgRNA, or a second guide

RNA, such as a second sgRNA, that hybridizes to a DNA sequence on the sense strand or the antisense strand that immediately precedes a Protospacer Adjacent Motif (PAM) sequence.

**[0528]** In some embodiments, the genome-modifying agent, e.g., Cas9, is targeted to the cleavage site on the sense strand by interacting with a first guide RNA, e.g., first sgRNA, that hybridizes to a sequence on the sense strand that immediately precedes a PAM sequence. In some embodiments, the genome-modifying agent, e.g., Cas9, is targeted to the cleavage site on the antisense strand by interacting with a second guide RNA, e.g., second sgRNA, that hybridizes to a sequence on the antisense strand that immediately precedes a PAM sequence.

**[0529]** In some embodiments, the first guide RNA, e.g., first sgNA, that is specific to the sense strand of a target gene of interest is used to target the recombinant nuclease, e.g., Cas9, to induce a SSB at a cleavage site within the sense strand of the target gene. In some embodiments, the first guide RNA, e.g., first sgNA, that is specific to the antisense strand of a target gene of interest is used to target the recombinant nuclease, e.g., Cas9, to induce a SSB at a cleavage site within the antisense strand of the target gene.

**[0530]** In some embodiments, the second guide RNA, e.g., second sgNA, that is specific to the sense strand of a target gene of interest used to target the recombinant nuclease, e.g., Cas9, to induce a SSB at a cleavage site within the sense strand of the target gene. In some embodiments, the second guide RNA, e.g., second sgNA, that is specific to the antisense strand of a target gene of interest is used to target the recombinant nuclease, e.g., Cas9, to induce a SSB at a cleavage site within the antisense strand of the target gene.

**[0531]** In some embodiments, the first guide RNA, e.g., first sgNA, that is specific to the sense strand of a target gene of interest is used to target the recombinant nuclease, e.g., Cas9, to induce a SSB at a cleavage site within the sense strand of the target gene; and the second guide RNA, e.g., second sgNA, that is specific to the antisense strand of a target gene of interest is used to target the recombinant nuclease, e.g., Cas9, to induce a SSB at a cleavage site within the antisense strand of the target gene.

**[0532]** In some embodiments, the first guide RNA, e.g., first sgNA, that is specific to the antisense strand of a target gene of interest is used to target the recombinant nuclease, e.g., Cas9, to induce a SSB at a cleavage site within the antisense strand of the target gene; and the second guide RNA, e.g., second sgNA, that is specific to the sense strand of a target gene of interest is used to target the recombinant nuclease, e.g., Cas9, to induce a SSB at a cleavage site within the sense strand of the target gene. In general, a guide RNA, e.g., a first guide RNA, such as a first sgRNA, or a second guide RNA, such as a second sgRNA, is any nucleotide sequence comprising a sequence, e.g., a crRNA sequence, that has sufficient complementarity with a target gene sequence to hybridize with the target gene sequence at the cleavage site and direct sequence-specific binding of the recombinant nuclease to a portion of the target gene that includes the cleavage site. Full complementarity (100%) is not necessarily required, so long as there is sufficient complementarity to cause hybridization and promote formation of a complex, e.g.,

CRISPR complex, that includes the recombinant nuclease, e.g., Cas9, and the guide RNA, e.g., the first guide RNA, such as the first sgRNA, or the second guide RNA, such as the second sgRNA.

**[0533]** In some embodiments, the cleavage site is situated at a site within the target gene that is homologous to a sequence comprised within the guide RNA, e.g., sgRNA. In some embodiments, the cleavage site of the sense strand is situated at a site within the sense strand of the target gene that is homologous to a sequence comprised within the first guide RNA, e.g., the first sgRNA. In some embodiments, the cleavage site of the antisense strand is situated at a site within the antisense strand of the target gene that is homologous to a sequence comprised within the first guide RNA, e.g., the first sgRNA. In some embodiments, the cleavage site of the sense strand is situated at a site within the sense strand of the target gene that is homologous to a sequence comprised within the second guide RNA, e.g., the second sgRNA. In some embodiments, the cleavage site of the antisense strand is situated at a site within the antisense strand of the target gene that is homologous to a sequence comprised within the second guide RNA, e.g., the second sgRNA. In some embodiments, the cleavage site of the sense strand is situated at a site within the sense strand of the target gene that is homologous to a sequence comprised within the first guide RNA, e.g., the first sgRNA; and the cleavage site of the antisense strand is situated at a site within the antisense strand of the target gene that is homologous to a sequence comprised within the second guide RNA, e.g., the second sgRNA. In some embodiments, the cleavage site of the antisense strand is situated at a site within the antisense strand of the target gene that is homologous to a sequence comprised within the first guide RNA, e.g., the first sgRNA; and the cleavage site of the sense strand is situated at a site within the sense strand of the target gene that is homologous to a sequence comprised within the second guide RNA, e.g., the second sgRNA. In some embodiments, the cleavage site of the antisense strand is situated at a site within the antisense strand of the target gene that is homologous to a sequence comprised within the second guide RNA, e.g., the second sgRNA; and the cleavage site of the sense strand is situated at a site within the sense strand of the target gene that is homologous to a sequence comprised within the first guide RNA, e.g., the first sgRNA.

**[0534]** In some embodiments, the sense strand comprises the targeting sequence, and the targeting sequence includes the SNP and a protospacer adjacent motif (PAM) sequence. In some embodiments, the sense strand comprises the targeting sequence, and the targeting sequence includes the SNP and a protospacer adjacent motif (PAM) sequence; and the antisense strand comprises a sequence that is complementary to the targeting sequence and includes a PAM sequence. In some embodiments, the antisense strand comprises the targeting sequence, and the targeting sequence includes the SNP and a protospacer adjacent motif (PAM) sequence. In some embodiments, the antisense strand comprises the targeting sequence, and the targeting sequence includes the SNP and a protospacer adjacent motif (PAM) sequence; and the sense strand comprises a sequence that is complementary to the targeting sequence and includes a PAM sequence.

**[0535]** In some embodiments, the cleavage site on the sense strand and/or the antisense strand is situated approximately 3 nucleotides upstream of the PAM sequence. In some embodiments, the cleavage site on the sense strand and/or the antisense strand is situated approximately 3 nucleotides upstream of the juncture between the guide RNA and the PAM sequence. In some embodiments, the cleavage site on the sense strand and/or the antisense strand is situated 3 nucleotides upstream of the PAM sequence. In some embodiments, the cleavage site on the sense strand and/or the antisense strand is situated 4 nucleotides upstream of the PAM sequence.

**[0536]** In some embodiments, the PAM sequence that is recognized by a recombinant nuclease is in the sense strand. In some embodiments, the PAM sequence that is recognized by a recombinant nuclease is in the antisense strand. In some embodiments, the PAM sequence that is recognized by a recombinant nuclease is in the sense strand and is in the antisense strand. In some embodiments, the PAM sequence on the sense strand and the PAM sequence on the antisense strand are outwardly facing. In some embodiments, the PAM sequence on the sense strand and the PAM sequence on the antisense strand comprise the same nucleic acid sequence, which can be any PAM sequence disclosed herein. In some embodiments, the PAM sequence on the sense strand and the PAM sequence on the antisense strand each comprise a different nucleic acid sequence, each of which can be any of the PAM sequences disclosed herein.

**[0537]** In some embodiments, the PAM sequence that is recognized by a recombinant nuclease, e.g., Cas9, differs depending on the particular recombinant nuclease and the bacterial species it is from

**[0538]** Methods for designing guide RNAs, e.g., sgRNAs, and their exemplary targeting sequences, e.g., crRNA sequences, can include those described in, e.g., International PCT Pub. Nos. WO2015/161276, WO2017/193107, and WO2017/093969. Exemplary guide RNA structures, including particular domains, are described in WO2015/161276, e.g., in FIGS. 1A-1G therein. Since guide RNA is an RNA molecule, it will comprise the base uracil (U), while any DNA encoding the guide RNA molecule will comprise the base thymine (T). In some embodiments, the guide RNA, e.g., sgRNA, comprises a CRISPR targeting RNA sequence (crRNA) and a trans-activating crRNA sequence (tracrRNA). In some embodiments, the first guide RNA, e.g., the first sgRNA, and the second guide RNA, e.g., the second sgRNA, each comprise a crRNA and a tracrRNA. In some embodiments, the guide RNA, e.g., sgRNA, is an RNA comprising, from 5' to 3': a crRNA sequence and a tracrRNA sequence. In some embodiments, each of the first guide RNA, e.g., first sgRNA, and the second guide RNA, e.g., second sgRNA, is an RNA comprising, from 5' to 3': a crRNA sequence and a tracrRNA sequence. In some embodiments, the crRNA and tracrRNA do not naturally occur together in the same sequence.

**[0539]** In some embodiments, the crRNA comprises a nucleotide sequence that is homologous, e.g., is at least 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% homologous, or is 100% homologous, to a portion of the target gene that includes the cleavage site. In some embodiments, the crRNA comprises a

nucleotide sequence that is 100% homologous to a portion of the target gene that includes the cleavage site. In some embodiments, the portion of the target gene that includes the cleavage site is a portion of the sense strand of the target gene that includes the cleavage site. In some embodiments, the portion of the target gene that includes the cleavage site is a portion of the antisense strand of the target gene that includes the cleavage site.

**[0540]** In some embodiments, the sgRNA comprises a crRNA sequence that is homologous to a sequence in the target gene that includes the cleavage site. In some embodiments, the first sgRNA comprises a crRNA sequence that is homologous to a sequence in the sense strand of the target gene that includes the cleavage site; and/or the second sgRNA comprises a crRNA sequence that is homologous to a sequence in the antisense strand of the target gene that includes the cleavage site. In some embodiments, the first sgRNA comprises a crRNA sequence that is homologous to a sequence in the antisense strand of the target gene that includes the cleavage site; and/or the second sgRNA comprises a crRNA sequence that is homologous to a sequence in the sense strand of the target gene that includes the cleavage site.

**[0541]** In some embodiments, the crRNA sequence has 100% sequence identity to a sequence in the target gene that includes the cleavage site. In some embodiments, the crRNA sequence of the first sgRNA has 100% sequence identity to a sequence in the sense strand of the target gene that includes the cleavage site; and/or the crRNA sequence of the second sgRNA has 100% sequence identity to a sequence in the antisense strand of the target gene that includes the cleavage site. In some embodiments, the crRNA sequence of the first sgRNA has 100% sequence identity to a sequence in the antisense strand of the target gene that includes the cleavage site; and/or the crRNA sequence of the second sgRNA has 100% sequence identity to a sequence in the sense strand of the target gene that includes the cleavage site.

**[0542]** Guidance on the selection of crRNA sequences can be found, e.g., in Fu Y et al., Nat Biotechnol 2014 (doi: 10.1038/nbt.2808) and Sternberg SH et al., Nature 2014 (doi: 10.1038/nature13011). Examples of the placement of crRNA sequences within the guide RNA, e.g., sgRNA, structure include those described in WO2015/161276, e.g., in FIGS. 1A-1G therein.

**[0543]** Reference to "the crRNA" is to be understood as also including reference to the crRNA of the first sgRNA and the crRNA of the second sgRNA, each independently. Thus, embodiments referring to "the crRNA" is to be understood as independently referring to embodiments of (i) the crRNA, (ii) the crRNA of the first sgRNA, and (iii) the crRNA of the second sgRNA. In some embodiments, the crRNA is 15-27 nucleotides in length, i.e., the crRNA is 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, or 27 nucleotides in length. In some embodiments, the crRNA is 18-22 nucleotides in length. In some embodiments, the crRNA is 19-21 nucleotides in length. In some embodiments, the crRNA is 20 nucleotides in length.

**[0544]** In some embodiments, the crRNA is homologous to a portion of a target gene that includes the cleavage site. In some embodiments, the crRNA is homologous to a portion of the sense strand of the target gene that includes the cleavage site. In some embodiments, the crRNA is homologous to a portion of the antisense strand of the target gene that includes the cleavage site. In some embodiments, the crRNA of the first sgRNA is homologous to a portion of the sense strand of the target gene that includes the cleavage site; and the crRNA of the second sgRNA is homologous to a portion of the antisense strand of the target gene that includes the cleavage site.

**[0545]** In some embodiments, the crRNA is homologous to a portion of the antisense strand of a target gene that includes the cleavage site. In some embodiments, the crRNA is homologous to a portion of the sense strand of the target gene that includes the cleavage site. In some embodiments, the crRNA of the first sgRNA is homologous to a portion of the antisense strand of the target gene that includes the cleavage site; and the crRNA of the second sgRNA is homologous to a portion of the sense strand of the target gene that includes the cleavage site.

**[0546]** In some embodiments, the crRNA is homologous to a portion of a target gene that includes the cleavage site, and is 15-27 nucleotides in length, i.e., the crRNA is 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, or 27 nucleotides in length. In some embodiments, the portion of the target gene that includes the cleavage site is on the sense strand. In some embodiments, the portion of the target gene that includes the cleavage site is on the antisense strand.

**[0547]** In some embodiments, the crRNA is homologous to a portion, i.e., sequence, in the sense strand or the antisense strand of the target gene that includes the cleavage site and is immediately upstream of the PAM sequence.

**[0548]** In some embodiments, the tracrRNA sequence may be or comprise any sequence for tracrRNA that is used in any CRISPR/Cas9 system known in the art. Reference to "the tracrRNA" is to be understood as also including reference to the tracrRNA of the first sgRNA and the tracrRNA of the second sgRNA, each independently. Thus, embodiments referring to "the tracrRNA" is to be understood as independently referring to embodiments of (i) the tracrRNA, (ii) the tracrRNA of the first sgRNA, and (iii) the tracrRNA of the second sgRNA. Exemplary CRISPR/Cas9 systems, sgRNA, crRNA, and tracrRNA, and their manufacturing process and use include those described in, e.g., International PCT Pub. Nos. WO2015/161276, WO2017/193107 and WO2017/093969, and those described in, e.g., U.S. Patent Application Publication Nos. 20150232882, 20150203872, 20150184139, 20150079681, 20150073041, 20150056705, 20150031134, 20150020223, 20140357530, 20140335620, 20140310830, 20140273234, 20140273232, 20140273231, 20140256046, 20140248702, 20140242700, 20140242699, 20140242664, 20140234972, 20140227787, 20140189896, 20140186958, 20140186919, 20140186843, 20140179770, 20140179006, 20140170753, 20140093913, and 20140080216.

**[0549]** In some embodiments, the heterologous protein is associated with base editing. Base editors (BEs) are typically fusions of a Cas ("CRISPR-associated") domain and a nucleobase modification

domain (e.g., a natural or evolved deaminase, such as a cytidine deaminase that include APOBEC1 ("apolipoprotein B mRNA editing enzyme, catalytic polypeptide 1"), CDA ("cytidine deaminase"), and AID ("activation-induced cytidine deaminase")) domains. In some cases, base editors may also include proteins or domains that alter cellular DNA repair processes to increase the efficiency and/or stability of the resulting single-nucleotide change.

**[0550]** In some aspects, currently available base editors include cytidine base editors (e.g., BE4) that convert target C•G to T•A and adenine base editors (e.g., ABE7.10) that convert target A•T to G•C. In some aspects, Cas9-targeted deamination was first demonstrated in connection with a Base Editor (BE) system designed to induce base changes without introducing double-strand DNA breaks. Further Rat deaminase APOBEC1 (rAPOBEC1) fused to deactivated Cas9 (dCas9) was used to successfully convert cytidines to thymidines upstream of the PAM of the sgRNA. In some aspects, this first BE system was optimized by changing the dCas9 to a "nickase" Cas9 D10A, which nicks the strand opposite the deaminated cytidine. Without being bound by theory, this is expected to initiate long-patch base excision repair (BER), where the deaminated strand is preferentially used to template the repair to produce a U:A base pair, which is then converted to T:A during DNA replication.

**[0551]** In some embodiments, the exogenous agent and/or heterologous protein is or encodes a base editor (e.g., a nucleobase editor). In some embodiments, the exogenous agent and/or heterologous protein is a nucleobase editor containing a first DNA binding protein domain that is catalytically inactive, a domain having base editing activity, and a second DNA binding protein domain having nickase activity, where the DNA binding protein domains are expressed on a single fusion protein or are expressed separately (e.g., on separate expression vectors). In some embodiments, the base editor is a fusion protein comprising a domain having base editing activity (e.g., cytidine deaminase or adenosine deaminase), and two nucleic acid programmable DNA binding protein domains (napDNAbp), a first comprising nickase activity and a second napDNAbp that is catalytically inactive, wherein at least the two napDNAbp are joined by a linker. In some embodiments, the base editor is a fusion protein that comprises a DNA domain of a CRISPR-Cas (e.g., Cas9) having nickase activity (nCas; nCas9), a catalytically inactive domain of a CRISPR-Cas protein (e.g., Cas9) having nucleic acid programmable DNA binding activity (dCas; e.g., dCas9), and a deaminase domain, wherein the dCas is joined to the nCas by a linker, and the dCas is immediately adjacent to the deaminase domain. In some embodiments, the base editor is a adenine-to-thymine or "ATBE" (or thymine-to-adenine or "TABE") transversion base editors. Exemplary base editor and base editor systems include any as described in patent publication Nos. US20220127622, US20210079366, US20200248169, US20210093667, US20210071163, WO2020181202, WO2021158921, WO2019126709, WO2020181178, WO2020181195, WO2020214842, WO2020181193, which are hereby incorporated in their entirety.

**[0552]** In some embodiments, the exogenous agent and/or heterologous protein is one for use in target-primed reverse transcription (TPRT) or "prime editing". In some embodiments, prime editing

mediates targeted insertions, deletions, all 12 possible base-to-base conversions, and combinations thereof in human cells without requiring DSBs or donor DNA templates.

**[0553]** Prime editing is a genome editing method that directly writes new genetic information into a specified DNA site using a nucleic acid programmable DNA binding protein ("napDNAbp") working in association with a polymerase (i.e., in the form of a fusion protein or otherwise provided in trans with the napDNAbp), wherein the prime editing system is programmed with a prime editing (PE) guide RNA ("PEgRNA") that both specifies the target site and templates the synthesis of the desired edit in the form of a replacement DNA strand by way of an extension (either DNA or RNA) engineered onto a guide RNA (e.g., at the 5′ or 3′ end, or at an internal portion of a guide RNA). The replacement strand containing the desired edit (e.g., a single nucleobase substitution) shares the same sequence as the endogenous strand of the target site to be edited (with the exception that it includes the desired edit). Through DNA repair and/or replication machinery, the endogenous strand of the target site is replaced by the newly synthesized replacement strand containing the desired edit. In some cases, prime editing may be thought of as a "search-and- replace" genome editing technology since the prime editors search and locate the desired target site to be edited, and encode a replacement strand containing a desired edit which is installed in place of the corresponding target site endogenous DNA strand at the same time. For example, prime editing can be adapted for conducting precision CRISPR/Cas-based genome editing in order to bypass double stranded breaks. In some embodiments, the heterologous protein is or encodes for a Cas protein-reverse transcriptase fusions or related systems to target a specific DNA sequence with a guide RNA, generate a single strand nick at the target site, and use the nicked DNA as a primer for reverse transcription of an engineered reverse transcriptase template that is integrated with the guide RNA. In some embodiments, the prime editor protein is paired with two prime editing guide RNAs (pegRNAs) that template the synthesis of complementary DNA flaps on opposing strands of genomic DNA, resulting in the replacement of endogenous DNA sequence between the PE-induced nick sites with pegRNA-encoded sequences.

**[0554]** In some embodiments, the exogenous agent and/or heterologous protein is or encodes for a primer editor that is a reverse transcriptase, or any DNA polymerase known in the art. Thus, in one aspect, the prime editor may comprise Cas9 (or an equivalent napDNAbp) which is programmed to target a DNA sequence by associating it with a specialized guide RNA (i.e., PEgRNA) containing a spacer sequence that anneals to a complementary protospacer in the target DNA. Such methods include any disclosed in Anzalone et al., (doi.org/10.1038/s41586-019-1711-4), or in PCT publication Nos. WO2020191248, WO2021226558, or WO2022067130, which are hereby incorporated in their entirety.

**[0555]** In some embodiments, the exogenous agent and/or heterologous protein is for use in Programmable Addition via Site-specific Targeting Elements (PASTE). In some aspects, PASTE is platform in which genomic insertion is directed via a CRISPR-Cas9 nickase fused to both a reverse transcriptase and serine integrase. As described in Ioannidi et al. (doi.org/10.1101/2021.11.01.466786),

PASTE does not generate double stranded breaks, but allowed for integration of sequences as large as ~36 kb. In some embodiments, the serine integrase can be any known in the art. In some embodiments, the serine integrase has sufficient orthogonality such that PASTE can be used for multiplexed gene integration, simultaneously integrating at least two different genes at least two genomic loci. In some embodiments, PASTE has editing efficiencies comparable to or better than those of homology directed repair or non-homologous end joining based integration, with activity in nondividing cells and fewer detectable off-target events.

**[0556]** In some embodiments, the exogenous agent and/or heterologous protein is or encodes one or more polypeptides having an activity selected from the group consisting of: nuclease activity (e.g., programmable nuclease activity); nickase activity (e.g., programmable nickase activity); homing activity (e.g., programmable DNA binding activity); nucleic acid polymerase activity (e.g., DNA polymerase or RNA polymerase activity); integrase activity; recombinase activity; or base editing activity (e.g., cytidine deaminase or adenosine deaminase activity).

**[0557]** In some embodiments, delivery of the nuclease is by a provided vector encoding the nuclease (e.g. Cas).

**[0558]** In some embodiments, the provided lipid particles contain a nuclease protein and the nuclease protein is directly delivered to a target cell. Methods of delivering a nuclease protein include those as described, for example, in Cai et al. Elife, 2014, 3:e01911 and International patent publication No. WO2017068077. For instance, provided lipid particles comprise one or more Cas protein(s), such as Cas9. In some embodiments, the nuclease protein (e.g. Cas, such as Cas 9) is engineered as a chimeric nuclease protein with a viral structural protein (e.g. GAG) for packaging into the lipid particle (e.g. lentiviral vector particle, VLP, or gesicle). For instance, a chimeric Cas9-protein fusion with the structural GAG protein can be packaged inside a lipid particle. In some embodiments, the fusion protein is a cleavable fusion protein between (i) a viral structural protein (e.g. GAG) and (ii) a nuclease protein (e.g. Cas protein, such as Cas9). In some embodiments, the fusion protein is a cleavable fusion protein between (i) a viral matrix (MA) protein and (ii) a nuclease protein (e.g. Cas protein, such as Cas9). In some embodiments, the particle contains a nuclease protein (e.g., Cas protein, such as Cas 9) immediately downstream of the gag start codon.

**[0559]** In some embodiments, the provided lipid particles contain mRNA encoding a Cas nuclease (e.g., Cas9). In some embodiments, the provided lipid particles contain guide RNA (gRNA), such as a single guide RNA (sgRNA).

**[0560]** In some embodiments, a dCas9 version of the CRISPR/Cas9 system can be used to target protein domains for transcriptional regulation, epigenetic modification, and microscopic visualization of specific genome loci.

**[0561]** In some embodiments, the provided virus particles (e.g. lentiviral particles) containing a Cas nuclease (e.g. Cas9) further comprise, or is further complexed with, one or more CRISPR-Cas system

guide RNA(s) for targeting a desired target gene. In some embodiments, the CRISPR guide RNAs are efficiently encapsulated in the CAS-containing viral particles. In some embodiments, the provided virus particles (e.g. lentiviral particles) further comprises, or is further complexed with a targeting nucleic acid.

#### *4. Small Molecules*

**[0562]** In some embodiments, the exogenous agent includes a small molecule, e.g., ions (e.g. $Ca^{2+}$, C1-, $Fe^{2+}$), carbohydrates, lipids, reactive oxygen species, reactive nitrogen species, isoprenoids, signaling molecules, heme, polypeptide cofactors, electron accepting compounds, electron donating compounds, metabolites, ligands, and any combination thereof. In some embodiments the small molecule is a pharmaceutical that interacts with a target in the cell. In some embodiments the small molecule targets a protein in the cell for degradation. In some embodiments the small molecule targets a protein in the cell for degradation by localizing the protein to the proteasome. In some embodiments that small molecule is a proteolysis targeting chimera molecule (PROTAC).

**[0563]** In some embodiments, the exogenous agent includes a mixture of proteins, nucleic acids, or metabolites, e.g., multiple polypeptides, multiple nucleic acids, multiple small molecules; combinations of nucleic acids, polypeptides, and small molecules; ribonucleoprotein complexes (e.g. Cas9-gRNA complex); multiple transcription factors, multiple epigenetic factors, reprogramming factors (e.g. Oct4, Sox2, cMyc, and Klf4); multiple regulatory RNAs; and any combination thereof.

## III. PHARMACEUTICAL COMPOSITIONS

**[0564]** Also provided are compositions containing the lipid particles herein, including pharmaceutical compositions and formulations. The pharmaceutical compositions can include any of the described lipid particles.

**[0565]** The present disclosure also provides, in some aspects, a pharmaceutical composition comprising the composition described herein and pharmaceutically acceptable carrier.

**[0566]** The term "pharmaceutical formulation" refers to a preparation which is in such form as to permit the biological activity of an active ingredient contained therein to be effective, and which contains no additional components which are unacceptably toxic to a subject to which the formulation would be administered.

**[0567]** A "pharmaceutically acceptable carrier" refers to an ingredient in a pharmaceutical formulation, other than an active ingredient, which is nontoxic to a subject. A pharmaceutically acceptable carrier includes, but is not limited to, a buffer, excipient, stabilizer, or preservative.

**[0568]** In some aspects, the choice of carrier is determined in part by the particular lipid particle and/or by the method of administration. Accordingly, there are a variety of suitable formulations. For example, the pharmaceutical composition can contain preservatives. Suitable preservatives may include, for example, methylparaben, propylparaben, sodium benzoate, and benzalkonium chloride. In some

aspects, a mixture of two or more preservatives is used. The preservative or mixtures thereof are typically present in an amount of about 0.0001% to about 2% by weight of the total composition. Carriers are described, e.g., by Remington's Pharmaceutical Sciences 16th edition, Osol, A. Ed. (1980). Pharmaceutically acceptable carriers are generally nontoxic to recipients at the dosages and concentrations employed, and include, but are not limited to: buffers such as phosphate, citrate, and other organic acids; antioxidants including ascorbic acid and methionine; preservatives (such as octadecyldimethylbenzyl ammonium chloride; hexamethonium chloride; benzalkonium chloride; benzethonium chloride; phenol, butyl or benzyl alcohol; alkyl parabens such as methyl or propyl paraben; catechol; resorcinol; cyclohexanol; 3-pentanol; and m-cresol); low molecular weight (less than about 10 residues) polypeptides; proteins, such as serum albumin, gelatin, or immunoglobulins; hydrophilic polymers such as polyvinylpyrrolidone; amino acids such as glycine, glutamine, asparagine, histidine, arginine, or lysine; monosaccharides, disaccharides, and other carbohydrates including glucose, mannose, or dextrins; chelating agents such as EDTA; sugars such as sucrose, mannitol, trehalose or sorbitol; salt-forming counter-ions such as sodium; metal complexes (e.g. Zn-protein complexes); and/or non-ionic surfactants such as polyethylene glycol (PEG).

**[0569]** In some embodiments, the lipid particle meets a pharmaceutical or good manufacturing practices (GMP) standard. In some embodiments, the lipid particle is made according to good manufacturing practices (GMP). In some embodiments, the lipid particle has a pathogen level below a predetermined reference value, e.g., is substantially free of pathogens. In some embodiments, the lipid particle has a contaminant level below a predetermined reference value, e.g., is substantially free of contaminants. In some embodiments, the lipid particle has low immunogenicity.

**[0570]** In some embodiments, formulations of the pharmaceutical compositions described herein may be prepared by any method known or hereafter developed in the art of pharmacology. In some embodiments, preparatory methods include the step of bringing the active ingredient into association with a carrier or one or more other accessory ingredients, and then, if necessary or desirable, shaping or packaging the product into a desired single- or multi-dose unit.

**[0571]** In some embodiments, a “unit dose” is a discrete amount of the pharmaceutical composition comprising a predetermined amount of the active ingredient. In some embodiments, the amount of the active ingredient is generally equal to the dosage of the active ingredient that would be administered to a subject or a convenient fraction of such a dosage such as, for example, one-half or one-third of such a dosage. In some embodiments, the unit dosage form may be for a single daily dose or one of multiple daily doses (e.g., about 1 to 4 or more times per day). In some embodiments, when multiple daily doses are used, the unit dosage form may be the same or different for each dose.

**[0572]** In some embodiments, the lipid particle containing the variant NiV-G is a viral vector or virus-like particle (e.g., Section II). In some embodiments, the compositions provided herein can be formulated in dosage units of genome copies (GC). Suitable method for determining GC have been

described and include, e.g., qPCR or digital droplet PCR (ddPCR) as described in, e.g., M. Lock et al, Hu Gene Therapy Methods, Hum Gene Ther Methods 25(2):115-25. 2014, which is incorporated herein by reference. In some embodiments, the dosage of administration of a viral vector or virus-like particle is from about $10^4$ to about $10^{10}$ GC units, inclusive. In some embodiments, the dosage of administration of a viral vector or virus-like particle is from about $10^9$ to about $10^{15}$ GC units, inclusive. In some embodiments, the dosage of administration of a viral vector or virus-like particle is from about $10^5$ to about $10^9$ GC units, inclusive. In some embodiments, the dosage of administration of a viral vector or virus-like particle is from about $10^6$ to about $10^9$ GC units, inclusive. In some embodiments, the dosage of administration of a viral vector or virus-like particle is from about $10^{12}$ to about $10^{14}$ GC units, inclusive. In some embodiments, the dosage of administration is $1.0\times10^9$ GC units, $5.0\times10^9$ GC units, $1.0\times10^{10}$ GC units, $5.0\times10^{10}$ GC units, $1.0\times10^{11}$ GC units, $5.0\times10^{11}$ GC units, $1.0\times10^{12}$ GC units, $5.0\times10^{12}$ GC units, or $1.0\times10^{13}$ GC units, $5.0\times10^{13}$ GC units, $1.0\times10^{14}$ GC units, $5.0\times10^{14}$ GC units, or $1.0\times10^{15}$ GC units.

**[0573]** In some embodiments, the dosage of administration of a viral vector or virus-like particle is from about $10^4$ to about $10^{10}$ infectious units, inclusive. In some embodiments, the dosage of administration of a viral vector or virus-like particle is from about $10^9$ to about $10^{15}$ infectious units, inclusive In some embodiments, the dosage of administration of a viral vector or virus-like particle is from about $10^5$ to about $10^9$ infectious units. In some embodiments, the dosage of administration of a viral vector or virus-like particle is from about $10^6$ to about $10^9$ infectious units. In some embodiments, the dosage of administration of a viral vector or virus-like particle is from about $10^{12}$ to about $10^{14}$ infectious units, inclusive. In some embodiments, the dosage of administration is $1.0\times10^9$ infectious units, $5.0\times10^9$ infectious units, $1.0\times10^{10}$ infectious units, $5.0\times10^{10}$ infectious units, $1.0\times10^{11}$ infectious units, 5.0×1011 infectious units, $1.0\times10^{12}$ infectious units, $5.0\times10^{12}$ infectious units, or $1.0\times10^{13}$ infectious units, $5.0\times10^{13}$ infectious units, $1.0\times10^{14}$ infectious units, $5.0\times10^{14}$ infectious units, or $1.0\times10^{15}$ infectious units. The techniques available for quantifying infectious units are routine in the art and include viral particle number determination, fluorescence microscopy, and titer by plaque assay. For example, the number of adenovirus particles can be determined by measuring the absorbance at A260. Similarly, infectious units can also be determined by quantitative immunofluorescence of vector specific proteins using monoclonal antibodies or by plaque assay.

**[0574]** In some embodiments, methods that calculate the infectious units include the plaque assay, in which titrations of the virus are grown on cell monolayers and the number of plaques is counted after several days to several weeks. For example, the infectious titer is determined, such as by plaque assay, for example an assay to assess cytopathic effects (CPE). In some embodiments, a CPE assay is performed by serially diluting virus on monolayers of cells, such as HFF cells, that are overlaid with agarose. After incubation for a time period to achieve a cytopathic effect, such as for about 3 to 28 days, generally 7 to 10 days, the cells can be fixed and foci of absent cells visualized as plaques are

determined. In some embodiments, infectious units can be determined using an endpoint dilution (TCID50) method, which determines the dilution of virus at which 50% of the cell cultures are infected and hence, generally, can determine the titer within a certain range, such as one log.

**[0575]** In some embodiments, the dosage of administration of a viral vector or virus-like particle is from about $10^4$ to about $10^{10}$ plaque forming units (pfu), inclusive. In some embodiments, the dosage of administration of a viral vector or virus-like particle is from about $10^9$ to about $10^{15}$ pfu, inclusive In some embodiments, the dosage of administration of a viral vector or virus-like particle is from about $10^5$ to about $10^9$ pfu. In some embodiments, the dosage of administration of a viral vector or virus-like particle is from about $10^6$ to about $10^9$ pfu. In some embodiments, the dosage of administration of a viral vector or virus-like particle is from about 1012 to about $10^{14}$ pfu, inclusive. In some embodiments, the dosage of administration is $1.0\times10^9$ pfu, $5.0\times10^9$ pfu, $1.0\times10^{10}$ pfu, $5.0\times10^{10}$ pfu, $1.0\times10^{11}$ pfu, $5.0\times10^{11}$ pfu, $1.0\times10^{12}$ pfu, $5.0\times10^{12}$ pfu, or $1.0\times10^{13}$ pfu, $5.0\times10^{13}$ pfu, $1.0\times10^{14}$ pfu, $5.0\times10^{14}$ pfu, or $1.0\times10^{15}$ pfu.

**[0576]** In some embodiments, the subject will receive a single injection. In some embodiments, administration can be repeated at daily/weekly/monthly intervals for an indefinite period and/or until the efficacy of the treatment has been established. As set forth herein, the efficacy of treatment can be determined by evaluating the symptoms and clinical parameters described herein and/or by detecting a desired response.

**[0577]** The exact amount of vehicle provided lipid particle required will vary from subject to subject, depending on the species, age, weight and general condition of the subject, the particular polynucleic acid, polypeptide, or vector used, its mode of administration etc. An appropriate amount can be determined by one of ordinary skill in the art using only routine experimentation given the teachings herein.

**[0578]** Compositions in some embodiments are provided as sterile liquid preparations, e.g., isotonic aqueous solutions, suspensions, emulsions, dispersions, or viscous compositions, which may in some aspects be buffered to a selected pH. Liquid preparations are normally easier to prepare than gels, other viscous compositions, and solid compositions. Additionally, liquid compositions are somewhat more convenient to administer, especially by injection. Viscous compositions, on the other hand, can be formulated within the appropriate viscosity range to provide longer contact periods with specific tissues. Liquid or viscous compositions can comprise carriers, which can be a solvent or dispersing medium containing, for example, water, saline, phosphate buffered saline, polyol (for example, glycerol, propylene glycol, liquid polyethylene glycol) and suitable mixtures thereof.

**[0579]** Sterile injectable solutions can be prepared by incorporating the lipid particles in a solvent, such as in admixture with a suitable carrier, diluent, or excipient such as sterile water, physiological saline, glucose, dextrose, or the like. The compositions can also be lyophilized. The compositions can contain auxiliary substances such as wetting, dispersing, or emulsifying agents (e.g., methylcellulose),

pH buffering agents, gelling or viscosity enhancing additives, preservatives, flavoring agents, colors, and the like, depending upon the route of administration and the preparation desired. Standard texts may in some aspects be consulted to prepare suitable preparations.

**[0580]** Injectables can be prepared in conventional forms, either as liquid solutions or suspensions, solid forms suitable for solution of suspension in liquid prior to injection, or as emulsions. As used herein, "parenteral administration" includes intradermal, intranasal, subcutaneous, intramuscular, intraperitoneal, intravenous and intratracheal routes, as well as a slow release or sustained release system such that a constant dosage is maintained.

**[0581]** Various additives which enhance the stability and sterility of the compositions, including antimicrobial preservatives, antioxidants, chelating agents, and buffers, can be added. Prevention of the action of microorganisms can be ensured by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, sorbic acid, and the like. Prolonged absorption of the injectable pharmaceutical form can be brought about by the use of agents delaying absorption, for example, aluminum monostearate and gelatin.

**[0582]** Sustained-release preparations may be prepared. Suitable examples of sustained-release preparations include semipermeable matrices of solid hydrophobic polymers containing the antibody, which matrices are in the form of shaped articles, e.g. films, or microcapsules.

**[0583]** In some embodiments, vehicle formulations may comprise cyroprotectants. As used herein, there term “cryoprotectant” refers to one or more agent that when combined with a given substance, helps to reduce or eliminate damage to that substance that occurs upon freezing. In some embodiments, cryoprotectants are combined with vector vehicles in order to stabilize them during freezing. In some aspects, Frozen storage of RNA between –20° C. and –80° C. may be advantageous for long term (e.g. 36 months) stability of polynucleotide. In some embodiments, the RNA species is mRNA. In some embodiments, cryoprotectants are included in vehicle formulations to stabilize polynucleotide through freeze/thaw cycles and under frozen storage conditions. Cryoprotectants of the provided embodiments may include, but are not limited to sucrose, trehalose, lactose, glycerol, dextrose, raffinose and/or mannitol. Trehalose is listed by the Food and Drug Administration as being generally regarded as safe (GRAS) and is commonly used in commercial pharmaceutical formulations.

**[0584]** The formulations to be used for in vivo administration are generally sterile. Sterility may be readily accomplished, e.g., by filtration through sterile filtration membranes.

## IV. KITS AND METHODS OF USE

**[0585]** In some embodiments, the lipid particles provided herein or pharmaceutical compositions containing same can be administered to a subject, e.g. a mammal, e.g. a human. In such embodiments, the subject may be at risk of, may have a symptom of, or may be diagnosed with or identified as having, a particular disease or condition. In one embodiment, the subject has cancer. In one embodiment, the

subject has an infectious disease. In some embodiments, the lipid particle contains nucleic acid sequences encoding an exogenous agent for treating the disease or condition in the subject. For example, the exogenous agent is one that targets or is specific for a protein of a neoplastic cells and the lipid particle is administered to a subject for treating a tumor or cancer in the subject. In another example, the exogenous agent is an inflammatory mediator or immune molecule, such as a cytokine, and lipid particle is administered to a subject for treating any condition in which it is desired to modulate (e.g. increase) the immune response, such as a cancer or infectious disease. In some embodiments, the lipid particle is administered in an effective amount or dose to effect treatment of the disease, condition or disorder. Provided herein are uses of any of the provided lipid particles in such methods and treatments, and in the preparation of a medicament in order to carry out such therapeutic methods. In some embodiments, the methods are carried out by administering the lipid particle or compositions comprising the same, to the subject having, having had, or suspected of having the disease or condition or disorder. In some embodiments, the methods thereby treat the disease or condition or disorder in the subject. Also provided herein are uses of any of the compositions, such as pharmaceutical compositions provided herein, for the treatment of a disease, condition or disorder associated with a particular gene or protein targeted by or provided by the exogenous agent.

**[0586]** In some embodiments, the provided methods or uses involve administration of a pharmaceutical composition comprising oral, inhaled, transdermal or parenteral (including intravenous, intratumoral, intraperitoneal, intramuscular, intracavity, intranodal and subcutaneous) administration. In some embodiments, the lipid particle may be administered alone or formulated as a pharmaceutical composition. In some embodiments, the lipid particle or compositions described herein can be administered to a subject, e.g., a mammal, e.g., a human. In some of any embodiments, the subject may be at risk of, may have a symptom of, or may be diagnosed with or identified as having, a particular disease or condition (e.g., a disease or condition described herein). In some embodiments, the disease is a disease or disorder.

**[0587]** In some embodiments, the lipid particles may be administered in the form of a unit-dose composition, such as a unit dose oral, parenteral, transdermal or inhaled composition. In some embodiments, the compositions are prepared by admixture and are adapted for oral, inhaled, transdermal or parenteral administration, and as such may be in the form of tablets, capsules, oral liquid preparations, powders, granules, lozenges, reconstitutable powders, injectable and infusible solutions or suspensions or suppositories or aerosols.

**[0588]** In some embodiments, the regimen of administration may affect what constitutes an effective amount. In some embodiments, the therapeutic formulations may be administered to the subject either prior to or after a diagnosis of disease. In some embodiments, several divided dosages, as well as staggered dosages may be administered daily or sequentially, or the dose may be continuously infused, or may be a bolus injection. In some embodiments, the dosages of the therapeutic formulations may be

proportionally increased or decreased as indicated by the exigencies of the therapeutic or prophylactic situation.

**[0589]** In some embodiments, the administration of the compositions of the present invention to a subject, preferably a mammal, more preferably a human, may be carried out using known procedures, at dosages and for periods of time effective to prevent or treat disease. In some embodiments, an effective amount of the therapeutic compound necessary to achieve a therapeutic effect may vary according to factors such as the activity of the particular compound employed; the time of administration; the rate of excretion of the compound; the duration of the treatment; other drugs, compounds or materials used in combination with the compound; the state of the disease or disorder, age, sex, weight, condition, general health and prior medical history of the subject being treated, and like factors well-known in the medical arts. In some embodiments, the dosage regimens may be adjusted to provide the optimum therapeutic response. In some embodiments, several divided doses may be administered daily or the dose may be proportionally reduced as indicated by the exigencies of the therapeutic situation. In some embodiments, the effective dose range for a therapeutic compound of the invention is from about 1 and 5,000 mg/kg of body weight/per day. One of ordinary skill in the art would be able to study the relevant factors and make the determination regarding the effective amount of the therapeutic compound without undue experimentation.

**[0590]** In some embodiments, the compound may be administered to a subject as frequently as several times daily, or it may be administered less frequently, such as once a day, once a week, once every two weeks, once a month, or even less frequently, such as once every several months or even once a year or less. In some embodiments, the amount of compound dosed per day may be administered, in non-limiting examples, every day, every other day, every 2 days, every 3 days, every 4 days, or every 5 days. In some embodiments, with every other day administration, a 5 mg per day dose may be initiated on Monday with a first subsequent 5 mg per day dose administered on Wednesday, a second subsequent 5 mg per day dose administered on Friday, and so on. The frequency of the dose will be readily apparent to the skilled artisan and will depend upon any number of factors, such as, but not limited to, the type and severity of the disease being treated, the type and age of the animal, etc.

**[0591]** In some embodiments, dosage levels of the active ingredients in the pharmaceutical compositions of this invention may be varied so as to obtain an amount of the active ingredient that is effective to achieve the desired therapeutic response for a particular subject, composition, and mode of administration, without being toxic to the subject.

**[0592]** A medical doctor, e.g., physician or veterinarian, having ordinary skill in the art may readily determine and prescribe the effective amount of the pharmaceutical composition required. In some embodiments, the physician or veterinarian could start doses of the compounds of the invention employed in the pharmaceutical composition at levels lower than that required in order to achieve the desired therapeutic effect and gradually increase the dosage until the desired effect is achieved.

**[0593]** In some embodiments, it is especially advantageous to formulate the compound in dosage unit form for ease of administration and uniformity of dosage. In some embodiments, dosage unit form as used herein refers to physically discrete units suited as unitary dosages for the subjects to be treated; each unit containing a predetermined quantity of therapeutic compound calculated to produce the desired therapeutic effect in association with the required pharmaceutical vehicle. In some embodiments, the dosage unit forms of the invention are dictated by and directly dependent on (a) the unique characteristics of the therapeutic compound and the particular therapeutic effect to be achieved, and (b) the limitations inherent in the art of compounding/formulating such a therapeutic compound for the treatment of a disease in a subject.

**[0594]** In some embodiments, the term "container" includes any receptacle for holding the pharmaceutical composition. In some embodiments, the container is the packaging that contains the pharmaceutical composition. In other embodiments, the container is not the packaging that contains the pharmaceutical composition, i.e., the container is a receptacle, such as a box or vial that contains the packaged pharmaceutical composition or unpackaged pharmaceutical composition and the instructions for use of the pharmaceutical composition. It should be understood that the instructions for use of the pharmaceutical composition may be contained on the packaging containing the pharmaceutical composition, and as such the instructions form an increased functional relationship to the packaged product. In some embodiments, instructions may contain information pertaining to the compound's ability to perform its intended function, e.g., treating or preventing a disease in a subject, or delivering an imaging or diagnostic agent to a subject.

**[0595]** In some embodiments, routes of administration of any of the compositions disclosed herein include oral, nasal, rectal, parenteral, sublingual, transdermal, transmucosal (e.g., sublingual, lingual, (trans)buccal, (trans)urethral, vaginal (e.g., trans- and perivaginally), (intra)nasal, and (trans)rectal), intravesical, intrapulmonary, intraduodenal, intragastrical, intrathecal, subcutaneous, intramuscular, intradermal, intra-arterial, intravenous, intrabronchial, inhalation, and topical administration.

**[0596]** In some of any embodiments, suitable compositions and dosage forms include, for example, tablets, capsules, caplets, pills, gel caps, troches, dispersions, suspensions, solutions, syrups, granules, beads, transdermal patches, gels, powders, pellets, magmas, lozenges, creams, pastes, plasters, lotions, discs, suppositories, liquid sprays for nasal or oral administration, dry powder or aerosolized formulations for inhalation, compositions and formulations for intravesical administration and the like.

**[0597]** In some embodiments, the lipid particle composition comprising an exogenous agent or cargo, may be used to deliver such exogenous agent or cargo to a cell tissue or subject. In some embodiments, delivery of a cargo by administration of a lipid particle composition described herein may modify cellular protein expression levels. In certain embodiments, the administered composition directs upregulation of (via expression in the cell, delivery in the cell, or induction within the cell) of one or more cargo (e.g., a polypeptide or mRNA) that provide a functional activity which is substantially absent

or reduced in the cell in which the polypeptide is delivered. In some embodiments, the missing functional activity may be enzymatic, structural, or regulatory in nature. In some embodiments, the administered composition directs up-regulation of one or more polypeptides that increases (e.g., synergistically) a functional activity which is present but substantially deficient in the cell in which the polypeptide is upregulated. In some of any embodiments, the administered composition directs downregulation of (via expression in the cell, delivery in the cell, or induction within the cell) of one or more cargo (e.g., a polypeptide, siRNA, or miRNA) that repress a functional activity which is present or upregulated in the cell in which the polypeptide, siRNA, or miRNA is delivered. In some of any embodiments, the upregulated functional activity may be enzymatic, structural, or regulatory in nature. In some embodiments, the administered composition directs down-regulation of one or more polypeptides that decreases (e.g., synergistically) a functional activity which is present or upregulated in the cell in which the polypeptide is downregulated. In some embodiments, the administered composition directs upregulation of certain functional activities and downregulation of other functional activities.

**[0598]** In some of any embodiments, the lipid particle composition (e.g., one comprising mitochondria or DNA) mediates an effect on a target cell, and the effect lasts for at least 1, 2, 3, 4, 5, 6, or 7 days, 2, 3, or 4 weeks, or 1, 2, 3, 6, or 12 months. In some embodiments (e.g., wherein the lipid particle composition comprises an exogenous protein), the effect lasts for less than 1, 2, 3, 4, 5, 6, or 7 days, 2, 3, or 4 weeks, or 1, 2, 3, 6, or 12 months.

**[0599]** In some of any embodiments, the lipid particle composition described herein is delivered ex-vivo to a cell or tissue, e.g., a human cell or tissue. In embodiments, the composition improves function of a cell or tissue ex-vivo, e.g., improves cell viability, respiration, or other function (e.g., another function described herein).

**[0600]** In some embodiments, the composition is delivered to an ex vivo tissue that is in an injured state (e.g., from trauma, disease, hypoxia, ischemia or other damage).

**[0601]** In some embodiments, the composition is delivered to an ex-vivo transplant (e.g., a tissue explant or tissue for transplantation, e.g., a human vein, a musculoskeletal graft such as bone or tendon, cornea, skin, heart valves, nerves; or an isolated or cultured organ, e.g., an organ to be transplanted into a human, e.g., a human heart, liver, lung, kidney, pancreas, intestine, thymus, eye). In some embodiments, the composition is delivered to the tissue or organ before, during and/or after transplantation.

**[0602]** In some embodiments, the composition is delivered, administered or contacted with a cell, e.g., a cell preparation. In some embodiments, the cell preparation may be a cell therapy preparation (a cell preparation intended for administration to a human subject). In embodiments, the cell preparation comprises cells expressing a chimeric antigen receptor (CAR), e.g., expressing a recombinant CAR. The cells expressing the CAR may be, e.g., T cells, Natural Killer (NK) cells, cytotoxic T lymphocytes (CTL), regulatory T cells. In embodiments, the cell preparation is a neural stem cell preparation. In embodiments, the cell preparation is a mesenchymal stem cell (MSC) preparation. In embodiments, the

cell preparation is a hematopoietic stem cell (HSC) preparation. In embodiments, the cell preparation is an islet cell preparation.

**[0603]** In some embodiments, the lipid particle compositions described herein can be administered to a subject, e.g., a mammal, e.g., a human. In such embodiments, the subject may be at risk of, may have a symptom of, or may be diagnosed with or identified as having, a particular disease or condition (e.g., a disease or condition described herein).

**[0604]** In some embodiments, the source of lipid particles are from the same subject that is administered a lipid particle composition. In other embodiments, they are different. In some embodiments, the source of lipid particles and recipient tissue may be autologous (from the same subject) or heterologous (from different subjects). In some embodiments, the donor tissue for lipid particle compositions described herein may be a different tissue type than the recipient tissue. In some embodiments, the donor tissue may be muscular tissue and the recipient tissue may be connective tissue (e.g., adipose tissue). In other embodiments, the donor tissue and recipient tissue may be of the same or different type, but from different organ systems.

**[0605]** In some embodiments, the lipid particle composition described herein may be administered to a subject having a cancer, an autoimmune disease, an infectious disease, a metabolic disease, a neurodegenerative disease, or a genetic disease (e.g., enzyme deficiency). In some embodiments, the subject is in need of regeneration.

**[0606]** In some embodiments, the lipid particle is co-administered with an inhibitor of a protein that inhibits membrane fusion. For example, Suppressyn is a human protein that inhibits cell-cell fusion (Sugimoto et al., "A novel human endogenous retroviral protein inhibits cell-cell fusion" Scientific Reports 3: 1462 (DOI: 10.1038/srep01462)). In some embodiments, the lipid particle particles is co-administered with an inhibitor of sypressyn, e.g., a siRNA or inhibitory antibody.

## V. EXEMPLARY EMBODIMENTS

**[0607]** Among the provided embodiments are:

1. A lipid particle, comprising:

(a) a paramyxovirus envelope attachment protein attached to a targeting moiety via a universal adapter system, said system comprising a tag polypeptide component and a catcher polypeptide component, wherein the paramyxovirus envelope attachment protein is fused to one of the components, and the targeting moiety is fused to the other of the components; and

(b) at least one paramyxovirus fusion (F) protein;

wherein the protein in (a) and (b) are exposed on the outside of the lipid bilayer.

2. The lipid particle of embodiment 1, wherein the tag polypeptide component and catcher polypeptide component are covalently bound.

3. A lipid particle, comprising:

(a) a paramyxovirus envelope attachment protein fused to a tag polypeptide component or fused to a catcher polypeptide component of a universal adapter system; and

(b) at least one paramyxovirus fusion (F) protein;

wherein the protein in (a) and (b) are exposed on the outside of the lipid bilayer.

4. The lipid particle of any of embodiments 1-3, wherein the paramyxovirus envelope attachment protein is fused to the tag polypeptide component of the universal adapter system.

5. The lipid particle of any of embodiments 1-3, wherein the paramyxovirus envelope attachment protein is fused to the catcher polypeptide component of the universal adapter system.

6. The lipid particle of any of embodiments 1-5, wherein the paramyxovirus envelope attachment protein is fused to the component of the universal adapter system at the C terminus of the paramyxovirus envelope attachment protein.

7. The lipid particle of any of embodiments 1-5, wherein the paramyxovirus envelope attachment protein is fused to the component of the universal adapter system by insertion of the component within a peptide loop of the paramyxovirus envelope attachment protein.

8. The lipid particle of embodiment 7, wherein the peptide loop of the paramyxovirus envelope attachment protein is selected from the group consisting of:

(i) amino acid residues 194-197;

(ii) amino acid residues 208-213;

(iii) amino acid residues 241-243;

(iv) amino acid residues 242-243;

(v) amino acid residues 388-391;

(vi) amino acid residues 402-405;

(vii) amino acid residues 489-491; and

(viii) amino acid residues 495-497;

optionally wherein the peptide loop is 388-391, 489-491, or 495-497.

9. The lipid particle of embodiment 5-8, wherein the fusion is direct or indirect via flexible linker, optionally a peptide linker.

10. The lipid particle of embodiment 9, wherein the peptide linker is a poly-Glycine-Serine (G4S) linker, optionally wherein the linker is set forth in any one of SEQ ID NOs. 7-10.

11. The lipid particle of any one of embodiments 1-10, wherein the universal adapter system comprises a tag component and a catcher component from a SpyTag/SpyCatcher system; a SnoopTag/SnoopCatcher system; a TEFCA tag/catcher system, or a DogTag/DogCatcher system.

12. The lipid particle of any one of embodiment 1-11, wherein the universal adapter system comprises a SpyTag tag component and a SpyCatcher catcher component selected from the group comprising:

(i) SpyTagΔ and SpyCatcherΔ (e.g., SEQ ID NOs 11 and 16); or

(ii) SpyTag003 and SpyCatcher003 (e.g., SEQ ID NOs 12 and 17).

13. The lipid particle of any one of embodiment 1-11, wherein the universal adapter system comprises:

(i) a SnoopTag tag component (e.g., SEQ ID NO. 13) and a SnoopCatcher catcher component (e.g., SEQ ID NO. 18);

(ii) a TEFCA tag component (e.g., SEQ ID NO. 14) and a TEFCACatcher catcher component (e.g., SEQ ID NO. 19); or

(iii) a DogTag tag component (e.g., SEQ ID NO. 15) and a DogCatcher catcher component (e.g., SEQ ID NO. 20).

14. The lipid particle of any of embodiments 1-2, and 4-13, wherein the targeting moiety is specific for a target molecule expressed on the surface of a target cell.

15. The lipid particle of embodiment 14, wherein the targeting moiety is selected from the group consisting of an antibody or antigen-binding fragment, an engineered binding domain, a nanobody, a DARPin, an Aptamer, an Affimer, an Affibody, a Knottin, an Avimer, a Monobody, an Anticalin, a Fynomer, and a targeting peptide.

16. The lipid particle of any of embodiments 14 or 15, wherein the targeting moiety is selected from the group consisting of a single domain antibody or a single chain variable fragment (scFv).

17. The lipid particle of embodiment 16, wherein the single domain antibody is a VHH.

18. The lipid particle of any of embodiments 1-17, wherein the paramyxovirus envelope attachment protein is an envelope attachment protein from a Nipah virus, Hendra virus, or Measles virus, or is a variant or biologically active portion thereof of any of the foregoing.

19. The lipid particle of embodiment 18, wherein the paramyxovirus envelope attachment protein is a variant Nipah G protein (NiV-G) that is a variant or a biologically active portion of a wild-type NiV-G.

20. The lipid particle of embodiment 19, wherein the variant NiV-G tis truncated by up to 40 contiguous amino acids at or near the N-terminus of the wild-type NiV-G set forth in SEQ ID NO:5.

21. The lipid particle of embodiment 19 or 20, wherein the variant NiV-G has a truncation of amino acids 2-34 of the wild-type NiV-G set forth in SEQ ID NO:5.

22. The lipid particle of any of embodiments 1-21, wherein the paramyxovirus envelope attachment protein comprises one or more mutations that reduces native tropism relative to the wild-type paramyxovirus envelope attachment protein not comprising the one or more mutations.

23. The lipid particle of any of embodiments 19-22, wherein the variant NiV-G exhibits reduced binding to Ephrin B2 or Ephrin B3.

24. The lipid particle of embodiment 13, wherein the variant NiV-G comprises: one or more amino acid substitutions corresponding to amino acid substitutions selected from the group consisting of E501A, W504A, Q530A and E533A with reference to numbering set forth in SEQ ID NO:5.

25. The lipid particle of embodiment 23 or embodiment 24, wherein the variant NiV-G comprises amino acid substitutions E501A, W504A, Q530A and E533A with reference to numbering set forth in SEQ ID NO:5.

26. The lipid particle of any of embodiments 19-25, wherein the variant NiV-G has the amino acid sequence set forth in SEQ ID NO: 6 or an amino acid sequence having at or about 80%, at least at or about 81 %, at least at or about 82%, at least at or about 83%, at or about 84%, at least at or about 85%, at least at or about 86%, or at least at or about 87%, at least at or about 88%, or at least at or about 89%, at least at or about 90%, at least at or about 91 %, at least at or about 92%, at least at or about 93%, at least at or about 94%, at least at or about 95%, at or about 96%, at least at or about 97%, at least at or about 98%, or at least at or about 99% sequence identity to SEQ ID NO:6.

27. The lipid particle of any of embodiments 19-26, wherein the variant NiV-G has the amino acid sequence set forth in SEQ ID NO:6.

28. The lipid particle of any of embodiments 1-27, wherein the at least one paramyxovirus fusion (F) protein is a wilt-type Nipah F protein (NiV-F) or is a functionally active variant or a biologically active portion thereof.

29. The lipid particle of embodiment 28, wherein the NiV-F protein or the biologically active portion is a truncated NiV-F that is truncated by at least 22 amino acids or at least 20 amino acids, at or near the C-terminus of wild-type NiV-F set forth in SEQ ID NO:1.

30. The lipid particle of any of embodiments 28 or 29, wherein the NiV-F protein or the biologically active portion thereof has a 22 amino acid truncation at or near the C-terminus of the wild-type NiV-F protein.

31. The lipid particle of any of embodiments 28-30, wherein the NiV-F protein comprises a deletion in it cytoplasmic tail and lacks amino acid residues 525-546 of SEQ ID NO:1.

32. The lipid particle of any of embodiments 28-31, wherein the NiV-F protein or the biologically active portion thereof has the sequence set forth in SEQ ID NO: 3 or 4, or a sequence of amino acids that exhibits at least at or about 80%, 85%, 90% or 95% sequence identity to the sequence set forth in SEQ ID NO: 3 or 4.

33. The lipid particle of any of embodiments 1-32, wherein the lipid particle is a vector, optionally a viral vector.

33. The lipid particle of embodiment 32, wherein the vector is a lentiviral vector.

34. The lipid particle of any of embodiments 1-33, wherein the lipid particle further comprises an exogenous agent for delivery to the target cell.

35. A combination, comprising:

(a) the lipid particle of any of embodiments 3-13, and 18-34, wherein the paramyxovirus envelope attachment protein is fused to a tag polypeptide component or a catcher polypeptide component of a universal adapter system, and

(b) a targeting moiety fused to the other component of the universal adapter system.

36. The combination of embodiment 35, wherein the universal adapter system comprises a tag component and a catcher component from a SpyTag/SpyCatcher system; a SnoopTag/SnoopCatcher system; a TEFCA tag/catcher system, or a DogTag/DogCatcher system.

37. The combination of embodiment 35 or 36, wherein the universal adapter system comprises a SpyTag tag component and a SpyCatcher catcher component selected from the group comprising:

(i) SpyTagΔ and SpyCatcherΔ (e.g., SEQ ID NOs 11 and 16); or

(ii) SpyTag003 and SpyCatcher003 (e.g., SEQ ID NOs 12 and 17).

38. The combination of any one of embodiments 35-37, wherein the universal adapter system comprises:

(i) a SnoopTag tag component (e.g., SEQ ID NO. 13) and a SnoopCatcher catcher component (e.g., SEQ ID NO. 18);

(ii) a TEFCA tag component (e.g., SEQ ID NO. 14) and a TEFCACatcher catcher component (e.g., SEQ ID NO. 19); or

(iii) a DogTag tag component (e.g., SEQ ID NO. 15) and a DogCatcher catcher component (e.g., SEQ ID NO. 20).

39. The combination of any of embodiments 35-38, wherein the targeting moiety is specific for a target molecule expressed on the surface of a target cell.

40. The combination of embodiment 39, wherein the targeting moiety is selected from the group consisting of an antibody or antigen-binding fragment, an engineered binding domain, a nanobody, a DARPin, an Aptamer, an Affimer, an Affibody, a Knottin, an Avimer, a Monobody, an Anticalin, a Fynomer, and a targeting peptide.

41. The combination of any of embodiments 39 or 40, wherein the targeting moiety is selected from the group consisting of a single domain antibody or a single chain variable fragment (scFv).

42. The combination of embodiment 41, wherein the single domain antibody is a VHH.

43. A method of making a lipid particle comprising a paramyxovirus envelope attachment protein attached to a targeting moiety via a universal adapter system, said system comprising a tag polypeptide component and a catcher polypeptide component, wherein the paramyxovirus envelope attachment protein is fused to one of the components, and the targeting moiety is fused to the other of the components; and at least one paramyxovirus fusion (F) protein, the method comprising:

(a) providing a cell that comprises the lipid particle of any one of embodiments 3-13 and 18-34;

(b) culturing the cell under condition that allow for production of the lipid particle,

(c) separating, enriching, or purifying the lipid particle from the cell, thereby making the lipid particle;

(d) mixing the lipid particle with a targeting moiety fused to the other component of the universal adapter system; and

(e) removing unbound targeting moiety from the lipid particles to collect lipid particles with the paramyxovirus envelope attachment protein attached to the targeting moiety.

44. The method of embodiment 43, further comprising concentrating the collected lipid particles.

45. A lipid particle produced by the method of embodiment 43 or 44.

46. A producer cell, comprising the lipid particle of any one of embodiments 3-13 and 18-34.

47. A kit comprising the combination of any of embodiment 35-42.

48. An article of manufacture comprising the kit of embodiment 47.

49. A method of selective transduction, the method comprising contacting a target cell with the lipid particle of any of embodiments 1, 2 and 4-34.

50. A method of selective transduction, the method comprising:

exposing a composition comprising a target cell with (1) the lipid particle of any of embodiments 3-13, and 18-34, wherein the lipid particle comprises a paramyxovirus envelope attachment protein fused to a tag polypeptide component or a catcher polypeptide component of a universal adapter system; and (2) a targeting moiety fused to the other component of the universal adapter system.

51. The method of embodiment 49 or embodiment 50 that is performed in vitro or ex vivo.

52. The method of any of embodiments 49 or embodiment 50 that is performed in vivo.

53. A method of selective transduction of a target cell in a subject, the method comprising administering the lipid particle of any of embodiments 1, 2 and 4-34 to a subject.

54. A method of selective transduction of a target cell in a subject, the method comprising:

(a) administering the lipid particle of any of embodiments 3-13, and 18-34 to a subject, wherein the lipid particle comprises a paramyxovirus envelope attachment protein fused to a tag polypeptide component or a catcher polypeptide component of a universal adapter system; and

(b) administering a targeting moiety fused to the other component of the universal adapter system to the subject.

55. The method of any of embodiments 50-52 and 54, wherein the universal adapter system comprises a tag component and a catcher component from a SpyTag/SpyCatcher system; a SnoopTag/SnoopCatcher system; a TEFCA tag/catcher system, or a DogTag/DogCatcher system.

56. The method of any of embodiments 50-52, 54 and 55, wherein the universal adapter system comprises a SpyTag tag component and a SpyCatcher catcher component selected from the group comprising:

(i) SpyTagΔ and SpyCatcherΔ (e.g., SEQ ID NOs 11 and 16); or

(ii) SpyTag003 and SpyCatcher003 (e.g., SEQ ID NOs 12 and 17).

57. The method of any of embodiments 50-52 and 54-56, wherein the universal adapter system comprises:

(i) a SnoopTag tag component (e.g., SEQ ID NO. 13) and a SnoopCatcher catcher component (e.g., SEQ ID NO. 18);

(ii) a TEFCA tag component (e.g., SEQ ID NO. 14) and a TEFCACatcher catcher component (e.g., SEQ ID NO. 19); or

(iii) a DogTag tag component (e.g., SEQ ID NO. 15) and a DogCatcher catcher component (e.g., SEQ ID NO. 20).

58. The method of any of embodiments 50-52 and 54-57, wherein the targeting moiety is specific for a target molecule expressed on the surface of the target cell.

59. The method of embodiment 58, wherein the targeting moiety is selected from the group consisting of an antibody or antigen-binding fragment, an engineered binding domain, a nanobody, a DARPin, an Aptamer, an Affimer, an Affibody, a Knottin, an Avimer, a Monobody, an Anticalin, a Fynomer, and a targeting peptide.

60. The method of any of embodiments 58 or 59, wherein the targeting moiety is selected from the group consisting of a single domain antibody or a single chain variable fragment (scFv).

61. The method of embodiment 60, wherein the single domain antibody is a VHH.

62. A paramyxovirus envelope attachment protein with a targeting moiety attached via a universal adapter system, said system comprising a tag polypeptide component and a catcher polypeptide component, wherein the paramyxovirus envelope attachment protein is fused to one of the components, and the targeting moiety is fused to the other of the components.

63. A paramyxovirus envelope attachment protein comprising one component of with a targeting moiety attached via a universal adapter system, said system comprising a tag polypeptide component and a catcher polypeptide component, wherein the paramyxovirus envelope attachment protein is fused to the one component.

64. The paramyxovirus envelope attachment protein of embodiment 62 or 63, wherein the component fused to the paramyxovirus envelope attachment protein is linked to the C-terminus of the paramyxovirus envelope attachment protein.

65. The paramyxovirus envelope attachment protein of embodiment 62 or 63, wherein the component fused to the paramyxovirus envelope attachment protein is inserted into a loop region of the paramyxovirus envelope attachment protein.

66. The paramyxovirus envelope attachment protein of embodiment 65, wherein the loop is selected from the group consisting of:

(i) amino acid residues 194-197;

(ii) amino acid residues 208-213;

(iii) amino acid residues 241-243;

(iv) amino acid residues 242-243;

(v) amino acid residues 388-391;

(vi) amino acid residues 402-405;

(vii) amino acid residues 489-491; and

(viii) amino acid residues 495-497;

optionally wherein the peptide loop is 388-391, 489-491, or 495-497.

67. The paramyxovirus envelope attachment protein of any of embodiments 62-66, wherein the paramyxovirus envelope attachment protein is from Nipah.

68. The paramyxovirus envelope attachment protein of embodiment 67, wherein the paramyxovirus envelope attachment protein is a NiV-G or is a biologically active variant or truncated form thereof.

69. The paramyxovirus envelope attachment protein of any of embodiments 62-68, wherein the targeting moiety is selected from the group consisting of an antibody or antigen-binding fragment, an engineered binding domain, a nanobody, a DARPin, an Aptamer, an Affimer, an Affibody, a Knottin, an Avimer, a Monobody, an Anticalin, a Fynomer, and a targeting peptide.

## VI. EXAMPLES

**[0608]** The following examples are included for illustrative purposes only and are not intended to limit the scope of the invention.

### Example 1 Expression of Lentiviral Vector with Adapters

**[0609]** A lentiviral vector was pseudotyped with a universal viral fusogen composed of an attachment glycoprotein G linked to a small peptide component of an adapter system, in which the peptide component is able to react with a cognate component coupled to a targeting agent (e.g., antibody or fragment) in an irreversible conjugation. The design can result in a modular system in which a binding agent (e.g., antibody or fragment) for retargeting the lentiviral vector to a target cell can be covalently added to the universal pseudotyped NiV-G fusogen after lentiviral production.

**[0610]** In an exemplary set of experiments, the adapter system known as SpyCatcher-SpyTag was used. As an adaptor-modified universal G protein, an exemplary G protein from Nipah virus (NiV-G) that contains an N-terminal truncation and is blinded to its native target was linked as a fusion to the tag or catcher of the system. In this experiment, either a SpyTagΔ (SEQ ID NO:11), SpyTag003 (SEQ ID NO:12) or SpyCatcherΔ (SEQ ID NO: 16) was linked as a fusion at the C-terminus or within a loop region of the NiV-G sequence GcΔ34 (Bender et al. 2016 PLoS Pathol 12(6):e1005641; set forth in SEQ ID NO:6).

**[0611]** To produce the lentiviral vector (LV), HEK293 producer cells were transfected with plasmids expressing viral vector proteins (gag/pol, rev) and a transfer plasmid encoding a transgene (e.g., enhanced green fluorescent protein (eGFP)). Envelope proteins were provided as plasmids expressing the adaptor-modified universal G protein described above and a Nipah F protein (NiV-F), in this case the exemplary NiV-F sequence NiV-FΔ22 (SEQ ID NO:3; or SEQ ID NO:4 without a signal sequence; Bender et al. 2016 PLoS). Following viral vector production, the cell culture was centrifuged to pellet

the cells and the supernatant containing crude virus was collected. As a control, lentiviral vectors also were engineered with a CD8-retargeted NiV-G fusogen in which an anti-CD8 scFv or VHH domain, instead of the adaptor protein, was linked by fusion to the NiV-G sequence GcΔ34.

**[0612]** The cognate tag or catcher was also engineered for expression as a fusion at either the C- or N- terminus, via a linker (e.g. (G4S)2; SEQ ID NO:9), to a binder. In this experiment, an exemplary scFv binder or a VHH binder, each targeted to CD8, was used. The sequences encoding the adaptor-modified binders were cloned into an expression vector and were transfected into Expi293 suspension cells. Approximately 4 days post-transfection, the adaptor-modified binder proteins were harvested.

**[0613]** Harvested lentiviral vector and binders as prepared above were then combined and co-incubated for approximately 3 hours at 0°C to allow association of cognate partners of the adapter system (e.g., SpyTag and SpyCatcher). Unbound binder-adapter complexes (e.g., unbound scFv or VHH-adapter complexes) were removed. Lentiviral vectors with the retargeted binder covalently added to the adaptor-modified G protein fusogen (CD8 retargeted universal LV) was collected after mixing or was further concentrated by either ultracentrifugation or using 100 kDa MWCO concentrator units. In these experiments, the CD8 retargeted universal LV had a lower concentration than the CD8 retargeted LV controls above; the CD8 controls were concentrated 2-3 times higher.

**[0614]** Initial experiments were conducted to assess the impact of insertion location, using SpyTagΔ as the universal adaptor at one of 28 insertion sites located in the NiV-G protein. Exemplary NiV-G constructs are shown in the **Table E1** below.

**Table E1**: Exemplary G Protein Constructs with Adapters

| Construct | NA Sequence (SEQ ID NO) | AA Sequence (SEQ ID NO) |
|---|---|---|
| NivG388_SpyTag_391_ | 513 | 515 |
| NivG.SpyTag | 514 | 516 |

**[0615]** Pseudotyped LVs were produced as described above by mixing with a cognate CD8 retargeted scFv or VHH binder fused to SpyCatcher retargeted to CD8. As shown for the retargeted scFv (**FIG. 1A**) or VHH (**FIG. 1B**), the adapter insertion sites on the G protein influence the titer, particularly for the scFv.

**[0616]** In a complementary experiment, CD8 retargeted universal LV's were generated as described above with an exemplary adapter at one of many insertion sites located in the NiV-G protein. Several of these insertion sites were observed to result in functional LV particles. As shown in **FIG. 1C**, comparison of two exemplary insertion sites, Insertion Site 1 and 2, showed that similar G protein expression and occupancy did result in differing potency (**FIG. 1D**).

**[0617]** In a further experiment, the following combination of adaptor-modified NiV-G pseudotyped lentiviral vectors and cognate-modified binders were generated for testing in which the universal adaptor component (SpyTagΔ, SpyTag003 or SpyCatcher003) was inserted into loop 388-391 of NiV-G):

| **Table E2: Exemplary adaptor-modified NiV-G pseudotyped lentiviral vectors** | | |
|---|---|---|
| **Name** | **Lentiviral Vector** | **Binder** |
| CD8_NiVG re-targeted LV Control | CD8 sdAb retargeted NiV-G | None |
| CD8_NiVG re-targeted LV Control | CD8 scFv retargeted NiV-G | None |
| CD8_SpyCatcherΔ-NivG388_SpyTagΔ LV | NiVG (388-391 loop)_SpyTagΔ | CD8 VHH_SpyCatcherΔ |
| CD8_SpyCatcherΔ-NivG388_SpyTag003 LV | NiVG (388-391 loop)_SpyTag003 | CD8 scFv_SpyCatcherΔ |
| CD8_SpyTag003-NivG388_SpyCatcher003 LV | NiVG (388-391 loop)_SpyCatcher003 | CD8 scFv_SpyTag003 |

**[0618]** Transduction efficiency of the produced pseudotyped CD8 retargeted universal LV or CD8 retargeted LV controls was assessed for GFP expression by flow cytometry, titer on transduced cells, as well as G protein expression. It was observed that pseudotyped vector retargeted covalently to the binder via the adapter system (CD8 retargeted universal LV) showed similar transduction efficiency to the control vector pseudotyped LV with a retargeted G-binder fusion protein, despite the control vector being concentrated approximately 2-3x higher (**FIG. 2**). Titers of CD8 retargeted universal LV were also observed to be similar or greater than control LV (**FIG. 3A**).

**[0619]** In a complementary set of experiments, CD8 retargeted universal LV were generated using a number of alternative exemplary coupling technologies, e.g., Universal Adapter Systems as described herein in Section I.A. Various configurations of SpyTag and SpyCatcher such as exemplified in table E1 were assessed. Viral vectors were screened for efficient delivery of a reporter GFP transgene *in vitro* as shown in **FIG. 3B**.

**[0620]** The expression of G protein on produced lentiviral vectors with the different adaptor systems in Table E1 was also assessed and is shown in **FIG. 4**. As shown, pseudotyped LVs with CD8 retargeted SpyTagΔ as the universal adaptor show similar G expression compared to G expression on control LV (**FIG. 4**).

**[0621]** Studies were carried out to assess the binder (e.g., scFv or VHH) occupancy on NiV-G when the adaptor-modified LV was mixed with increasing amounts of the cognate adaptor-modified binder. Occupancy was assessed by Western blot to distinguish the proportion of NiV-G compared with an exemplary NiV-G + binder-Catcher. In this experiment, the binder was also fused to a His tag for specific detection with an anti-His antibody, whereas the NiV-G was assessed using an anti-NiV-G antibody. LV particles were lysed, run on 4-20% Tris-Glycine gel and transferred onto PVDF membrane for detection.

As shown in **FIG. 5A**, it was observed that a greater quantity of total NiV-G was expressed as complexed NiV-G + scFv-Catcher (110 kDa) as a function of increasing binder concentration added to the lentiviral vector. These data are quantified in **FIG. 5B**, wherein it was observed that upwards of 60% of total NiV-G were coupled with a binder.

**[0622]** Transduction efficiency of the produced pseudotyped lentiviral vector with binder (i.e., vector with binder comprising one of two exemplary VHH constructs or one of four exemplary scFv constructs) was assessed for titer on transduced cells (**FIG. 5C** ). Measurable titers were achieved for all the binder formats tested, and it was observed that NiV-G occupancy between 20-60%, including between 30%-40%, allowed for highest titer among the various formats. A similar experiment was conducted with exemplary binders 1, 2, and 3 in **FIG 5D**, wherein a positive correlation was observed between G occupancy and titer (TU/mL), up to 40% G occupancy.

**[0623]** Different produced pseudotyped CD8 retargeted universal LV, in which the SpyTag was inserted into different loops of NiV-G and covalently attached to a binder by mixing with a cognate CD8 VHH_SpyCatcher, was assessed for titer on transduced cells similarly as described above. Measurable titers were achieved for all the loop insertion sites, although titer was observed to vary across the insertion sites (**FIG. 6A**). Titer as a function of G protein occupancy was also assessed as shown in **FIG. 6B**, wherein it was observed that G protein binder occupancy does impact lentiviral titer such that occupancy and titer were determined to be positively correlated **(FIG. 6C)**.

**[0624]** In a similar experiment, transduction efficiency and selective transduction of the produced pseudotyped retargeted universal LV was assessed *in vitro*. Universal LV were generated as described above and retargeted using exemplary adapters comprising one of two binders specific for exemplary receptors A (ASGR1) and/or B (CD8), respectively. Subsequently, two different HEK 293 cell lines which overexpress either receptor A or B (and the mixed population) were transduced. The evaluation of transduction and receptor A expression was done by flow cytometry 4 days after infection. Transduction was observed specifically in cells overexpressing the targeted receptor in **FIG. 8**.

**[0625]** Taken together, these results support the production of pseudotyped lentiviral vector using an adapter system that allows for similar and/or improved transduction, titer, and expression as an adapter-less vector.

## Example 2 Comparative Analysis of a Lentiviral Vector with Adapters and a Pseudotyped Lentiviral Vector

**[0626]** Similarly, as described above, a lentiviral vector was pseudotyped with a universal viral fusogen composed of an attachment glycoprotein G linked to a small peptide component of an adapter system (e.g., a lentiviral vector with an adapter). In an exemplary set of experiments, such a lentiviral vector expressing a G protein with an adapter ($G_{ADAPTER}$) was assessed against a similar vector expressing

a wildtype viral attachment protein ($G_{WT}$), or against a re-targeted attachment protein comprising a targeting moiety ($G_{RETARGETED}$) for factors such as expression, potency, titer, and transduction.

**[0627]** Lentiviral vectors (LV) were produced substantially as described above. HEK293 producer cells were transfected with plasmids expressing viral vector proteins (gag/pol, rev) and a transfer plasmid encoding a transgene (e.g., enhanced green fluorescent protein (eGFP)). Envelope proteins were provided as plasmids expressing the adaptor-modified universal G protein described above ($G_{ADAPTER}$) and a Nipah F protein (NiV-F), in this case the exemplary NiV-F sequence NiV-FΔ22 (SEQ ID NO:3; or SEQ ID NO:4 without a signal sequence; Bender et al. 2016 PLoS). For the wildtype and retargeted G controls, envelope proteins were provided as plasmids expressing the wildtype G protein ($G_{WT}$) or a retargeted G protein ($G_{RETARGETED}$) and a Nipah F protein (NiV-F), in this case the exemplary NiV-F sequence NiV-FΔ22 (SEQ ID NO:3; or SEQ ID NO:4 without a signal sequence; Bender et al. 2016 PLoS).

**[0628]** Following viral vector production, the cell culture was centrifuged to pellet the cells and the supernatant containing crude virus was collected.

**[0629]** In the case of the adapter-modified G ($G_{ADAPTER}$), the cognate tag or catcher was also engineered for expression as a fusion at either the C- or N- terminus, via a linker (e.g. (G4S)2; SEQ ID NO:9), to a binder. In this experiment, an exemplary scFv binder or a VHH binder, each targeted to CD8, was used. The sequences encoding the adaptor-modified binders were cloned into an expression vector and were transfected into Expi293 suspension cells. Approximately 4 days post-transfection, the adaptor-modified binder proteins were harvested.

**[0630]** Harvested lentiviral vector and binders as prepared above were then combined in an exemplary protocol, P3. Here, vector and binders were co-incubated for approximately 3 hours at 0°C to allow association of cognate partners of the adapter system (e.g., SpyTag and SpyCatcher). Unbound binder-adapter complexes (e.g., unbound scFv or VHH-adapter complexes) were removed. Lentiviral vectors with the retargeted binder covalently added to the adaptor-modified G protein fusogen ($G_{ADAPTER}$), as well as the control vectors ($G_{WT}$ and $G_{RETARGETED}$), were collected after mixing or was further concentrated by either ultracentrifugation or using 100 kDa MWCO concentrator units. Alternative preparation methods for lentiviral vector and binder production are shown in **FIG. 9**.

**[0631]** Initial experiments were conducted to verify the amount of G protein per lentiviral particle. Both Western blot and ExoView R100 were utilized. Generally Western blot can provide bulk expression data of the G protein, while ExoView R100 can reflect the range of G expression per particle. As shown in **FIG. 10,** the adapter-modified G ($G_{ADAPTER}$) was observed to have similar expression and incorporation of the G protein into the viral particles as that seen for viral vectors pseudotyped with $G_{WT}$.

**[0632]** Potency was also assessed as a function of titer normalized to the LV with $G_{WT}$. Lentiviral vector engineered with two alternative adapter-modified Gs ($G_{ADAPTER1}$+ $G_{ADAPTER2}$) as well as the $G_{WT}$ control vector were generated and screened for delivery of a reporter GFP transgene in vitro using

HEK293 cells overexpressing membrane-presented complementary adapter proteins As shown in **FIG. 11A** similar titers were seen for the $G_{ADAPTER}$ as with the $G_{WT}$ control vector.

**[0633]** The produced pseudotyped CD8 retargeted universal LV ($G_{ADAPTER}$) was also assessed for titer on SupT1 cells and then compared to CD8 retargeted LV produced with a single G-binder fusion protein ($G_{RETARGETED}$). As shown in **FIG. 11B**, the pseudotyped CD8 retargeted universal LV ($G_{ADAPTER}$) showed similar and/or higher potency than $G_{RETARGETED}$ LV. For instance, at least two exemplary retargeted universal LV ($G_{ADAPTER}$) displayed titers 18x or 50x greater than that observed with the single G-binder fusion protein ($G_{RETARGETED}$). These results suggest retargeted universal LV ($G_{ADAPTER}$) may rescue and improve potency of suboptimal binders.

**[0634]** Together, these experiments demonstrate that this retargeted universal LV coupling approach (e.g., a Universal Adapter System) maintains transduction selectivity of the targeted cells and we can rescue the suboptimal infectivity observed for several binders. The level of expression of the G protein and the resulting viral potency is comparable to that seen for viral vectors pseudotyped with $G_{WT}$.

## Example 3 Comparative Analysis of a Lentiviral Vector with Adapters and a Pseudotyped Lentiviral Vector

**[0635]** In an exemplary set of experiments similar to those performed in Example 1, the adapter system known as DogTag/DogCatcher was used. As an adaptor-modified universal G protein, an exemplary G protein from Nipah virus (NiV-G) that contains an N-terminal truncation and is blinded to its native target was linked as a fusion to the tag or catcher of the system. In this experiment, either a DogTag (SEQ ID NO: 511) was linked as a fusion at the C-terminus or within a loop region of the NiV-G sequence GcΔ34 (Bender et al. 2016 PLoS Pathol 12(6):e1005641; set forth in SEQ ID NO:6).

**[0636]** To produce the lentiviral vector (LV), HEK293 producer cells were transfected with plasmids expressing viral vector proteins (gag/pol, rev) and a transfer plasmid encoding a transgene (e.g., enhanced green fluorescent protein (eGFP)). Envelope proteins were provided as plasmids expressing the adaptor-modified universal G protein described above and a Nipah F protein (NiV-F), in this case the exemplary NiV-F sequence NiV-FΔ22 (SEQ ID NO:3; or SEQ ID NO:4 without a signal sequence; Bender et al. 2016 PLoS). Following viral vector production, the cell culture was centrifuged to pellet the cells and the supernatant containing crude virus was collected. As a control, lentiviral vectors also were engineered with a CD8-retargeted NiV-G fusogen in which an anti-CD8 scFv or VHH domain, instead of the adaptor protein, was linked by fusion to the NiV-G sequence GcΔ34.

**[0637]** The cognate tag (e.g., DogTag) or catcher (e.g., DogCatcher (SEQ ID NO: 512) was also engineered for expression as a fusion at either the C- or N- terminus, via a linker (e.g. (G4S)2; SEQ ID NO:9), to a binder. In this experiment, an exemplary scFv binder or a VHH binder, each targeted to CD8, was used. The sequences encoding the adaptor-modified binders were cloned into an expression vector

and were transfected into Expi293 suspension cells. Approximately 4 days post-transfection, the adaptor-modified binder proteins were harvested.

**[0638]** Harvested lentiviral vector and binders as prepared above were then combined and co-incubated for approximately 3 hours at 0°C to allow association of cognate partners of the adapter system (e.g., DogTag and DogCatcher). Unbound binder-adapter complexes (e.g., unbound scFv or VHH-adapter complexes) were removed. Lentiviral vectors with the retargeted binder covalently added to the adaptor-modified G protein fusogen (CD8 retargeted universal LV) was collected after mixing or was further concentrated by either ultracentrifugation or using 100 kDa MWCO concentrator units.

**[0639]** Different produced pseudotyped CD8 retargeted universal LV, in which the DogTag was inserted into different loops of NiV-G and covalently attached to a binder by mixing with a cognate CD8 VHH_DogCatcher, was assessed for titer on transduced cells similarly as described above. Measurable titers were achieved for all the loop insertion sites, although titer was observed to vary across the insertion sites (**FIG. 12**).

**[0640]** These results support the production of pseudotyped lentiviral vector using an adapter system that allows for similar and/or improved transduction, titer, and expression as an adapter-less vector.

**[0641]** The present invention is not intended to be limited in scope to the particular disclosed embodiments, which are provided, for example, to illustrate various aspects of the invention. Various modifications to the compositions and methods described will become apparent from the description and teachings herein. Such variations may be practiced without departing from the true scope and spirit of the disclosure and are intended to fall within the scope of the present disclosure.

## VII. SEQUENCES

| | | |
|---|---|---|
| 1 | MVVILDKRCYCNLLILILMISECSVGILHYEKLSKIGLVKGVTRKYK<br>IKSNPLTKDIVIKMIPNVSNMSQCTGSVMENYKTRLNGILTPIKGAL<br>EIYKNNTHDLVGDVRLAGVIMAGVAIGIATAAQITAGVALYEAMK<br>NADNINKLKSSIESTNEAVVKLQETAEKTVYVLTALQDYINTNLVP<br>TIDKISCKQTELSLDLALSKYLSDLLFVFGPNLQDPVSNSMTIQAIS<br>QAFGGNYETLLRTLGYATEDFDDLLESDSITGQIIYVDLSSYYIIVR<br>VYFPILTEIQQAYIQELLPVSFNNDNSEWISIVPNFILVRNTLISNIEIG<br>FCLITKRSVICNQDYATPMTNNMRECLTGSTEKCPRELVVSSHVPR<br>FALSNGVLFANCISVTCQCQTTGRAISQSGEQTLLMIDNTTCPTAV<br>LGNVIISLGKYLGSVNYNSEGIAIGPPVFTDKVDISSQISSMNQSLQ<br>QSKDYIKEAQRLLDTVNPSLISMLSMIILYVLSIASLCIGLITFISFIIV<br>EKKRNTYSRLEDRRVRPTSSGDLYYIGT | Nipah F0 |
| 2 | ILHYEKLSKIGLVKGVTRKYKIKSNPLTKDIVIKMIPNVSNMSQCTG<br>SVMENYKTRLNGILTPIKGALEIYKNNTHDLVGDVRLAGVIMAGV<br>AIGIATAAQITAGVALYEAMKNADNINKLKSSIESTNEAVVKLQET<br>AEKTVYVLTALQDYINTNLVPTIDKISCKQTELSLDLALSKYLSDL<br>LFVFGPNLQDPVSNSMTIQAISQAFGGNYETLLRTLGYATEDFDDL<br>LESDSITGQIIYVDLSSYYIIVRVYFPILTEIQQAYIQELLPVSFNNDN<br>SEWISIVPNFILVRNTLISNIEIGFCLITKRSVICNQDYATPMTNNMR<br>ECLTGSTEKCPRELVVSSHVPRFALSNGVLFANCISVTCQCQTTGR<br>AISQSGEQTLLMIDNTTCPTAVLGNVIISLGKYLGSVNYNSEGIAIG | Nipah virus F Protein, without signal sequence |

| | | |
|---|---|---|
| | PPVFTDKVDISSQISSMNQSLQQSKDYIKEAQRLLDTVNPSLISMLS MIILYVLSIASLCIGLITFISFIIVEKKRNTYSRLEDRRVRPTSSGDLY YIGT | |
| 3 | MVVILDKRCY CNLLILILMI SECSVGILHY EKLSKIGLVKGVTRKYKIKS NPLTKDIVIK MIPNVSNMSQCTGSVMENYK TRLNGILTPI KGALEIYKNN THDLVGDVRL AGVIMAGVAI GIATAAQITA GVALYEAMKN NINKLKSS IESTNEAVVK LQETAEKTVY VLTALQDYIN TNLVPTIDKI SCKQTELSLD LALSKYLSDL LFVFGPNLQD PVSNSMTIQA ISQAFGGNYE LLRTLGYAT EDFDDLLESD SITGQIIYVD LSSYYIIVRV YFPILTEIQQ AYIQELLPVS FNNDNSEWIS IVPNFILVRN TLISNIEIGF CLITKRSVIC NQDYATPMTN NMRECLTGST EKCPRELVVS SHVPRFALSN GVLFANCISV CQCQTTGRA ISQSGEQTLL MIDNTTCPTA VLGNVIISLG KYLGSVNYNS EGIAIGPPVF TDKVDISSQI SSMNQSLQQS KDYIKEAQRL LDTVNPSLIS MLSMIILYVL SIASLCIGLI TFISFIIVEK KRNT | FcDelta22 with signal sequence |
| 4 | ILHYEKLSKIGLVKGVTRKYKIKSNPLTKDIVIKMIPNVSNMSQCTG SVMENYKTRLNGILTPIKGALEIYKNNTHDLVGDVRLAGVIMAGV AIGIATAAQITAGVALYEAMKNADNINKLKSSIESTNEAVVKLQET AEKTVYVLTALQDYINTNLVPTIDKISCKQTELSLDLALSKYLSDL LFVFGPNLQDPVSNSMTIQA ISQAFGGNYETLLRTLGYAT EDFDDLLESD SITGQIIYVD LSSYYIIVRV YFPILTEIQQAYIQELLPVSFNNDNSEWISIVPNFILVRNTLISNIEIGF CLITKRSVIC NQDYATPMTN NMRECLTGST EKCPRELVVS SHVPRFALSN GVLFANCISV TCQCQTTGRA ISQSGEQTLL MIDNTTCPTA VLGNVIISLG KYLGSVNYNS EGIAIGPPVF TDKVDISSQI SSMNQSLQQS KDYIKEAQRL LDTVNPSLIS MLSMIILYVL SIASLCIGLI TFISFIIVEK KRNT | FcDelta22 |
| 5 | MPAENKKVRFENTTSDKGKIPSKVIKSYYGTMDIKKINEGLLDSKI LSAFNTVIALLGSIVIIVMNIMIIQNYTRSTDNQAVIKDALQGIQQQI KGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSKISQSTASINENVN EKCKFTLPPLKIHECNISCPNPLPFREYRPQTEGVSNLVGLPNNICL QKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMDEGYFAYSHLERI GSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVWTPPNPNTVYHC SAVYNNEFYYVLCAVSTVGDPILNSTYWSGSLMMTRLAVKPKSN GGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQGDTLYFPAVGFLV RTEFKYNDSNCPITKCQYSKPENCRLSMGIRPNSHYILRSGLLKYN LSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVFYQASFSWDTMI KFGDVLTVNPLVVNWRNNTVISRPGQSQCPRFNTCPEICWEGVYN DAFLIDRINWISAGVFLDSNQTAENPVFTVFKDNEILYRAQLASED TNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPKLFAVKIPEQCT | Nipah G protein (Uniprot Q9IH62) |
| 6 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQNYTRSTDNQA VIKDALQG IQQQIKGLAD KIGTEIGPKVSLIDTSSTITIPANIGLLGSKISQSTASIN ENVNEKCKFTLPPLKIHECN ISCPNPLPFR EYRPQTEGVS NLVGLPNNIC LQKTSNQILK PKLISYTLPV VGQSGTCITD PLLAMDEGYF AYSHLERIGS CSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVWTPPNPNTVYHCSA VYNNE FYYVLCAVSTVGDPILNSTYWSGSLMMTRLAVKPKSNGGGYNQH | (E501 A, W504A, Q530A, E533A) NiV G protein (Gc Δ 34) |

| | | |
|---|---|---|
| | QLALRSIEKGRYDKVMPYGPSGIKQGDTLYFPAVGFLVRTEFKYN DSNCPITKCQYSKPENCRLSMGIRPNSHYILRSGLLKYNLSDGENP KVVFI EISDQRLSIG SPSKIYDSLGQPVFYQASFSWDTMIKFGDVLTVNPLVVNW RNNTVISRPG QSQCPRFNTCPAICAEGVYN DAFLIDRINW ISAGVFLDSN ATAANPVFTV FKDNEILYRA QLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIR PKLFAVKIPE QC | |
| 7 | GGGGS | Linker |
| 8 | GGGGGS | Linker |
| 9 | (GGGGS)n | Linker |
| 10 | (GGGGGS)n | Linker |
| 11 | AHIVMVDAYKPTK | SpyTagΔ |
| 12 | RGVPHIVMVDAYKRYK | SpyTag003 |
| 13 | KLGDIEFIKVNK | SnoopTag |
| 14 | TEFCA | TEFCA Tag |
| 15 | DIPATYEFTDGKHYITNEPIPPK | DogTag |
| 16 | DSATHIKFSKRDEDGKELAGATMELRDSSGKTISTWISDGQVKDF YLYPGKYTFVETAAPDGYEVATAITFTVNEQGQVTVNG | SpyCatcherΔ |
| 17 | VTTLSGLSGEQGPSGDMTTEEDSATHIKFSKRDEDGRELAGATME LRDSSGKTISTWISDGHVKDFYLYPGKYTFVETAAPDGYEVATPIE FTVNEDGQVTVDGEATEGDAHT | SpyCatcher003 |
| 18 | KPLRGAVFSLQKQHPDYPDIYGAIDQNGTYQNVRTGEDGKLTFK NLSDGKYRLFENSEPAGYKPVQNKPIVAFQIVNGEVRDVTSIVPQ DIPATYEFTNGKHYITNEPIPPK | SnoopCatcher |
| 19 | AGELIHMVTLDKTGKKSFGICIVRGEVKDSPNTKTTGIFIKGIVPD SPAHLCGRLKVGDRILSLNGKDVRNSTEQAVIDLIKEADFKIELEI QTFDK | TEFCACatcher |
| 20 | KLGEIEFIKVDKTDKKPLRGAVFSLQKQHPDYPDIYGAIDQNGTY QDVRTGEDGKLTFTNLSDGKYRLIENSEPPGYKPVQNKPIVSFRIV DGEVRDVTSIVPQ | DogCatcher |
| 21 | NLVGLPNNICLQKTSNQILKPKLI | site 1 |
| 22 | PVVGQSGT | site 2 |
| 23 | MDEG | site 3 |
| 24 | IGSCSRGV | site 4 |
| 25 | DRGDEVP | site 5 |
| 26 | WTPPNPNTV | site 6 |
| 27 | NNEF | site 7 |
| 28 | VSTVGDPILNSTYWSGSL | site 8 |
| 29 | AVKPKSNG | site 9 |
| 30 | LALRSIEKGRYDK | site 10 |
| 31 | QGDT | site 11 |
| 32 | VRTEFKYNDSNCPITKCQYSKP | site 12 |
| 33 | MGIRPNSHY | site 13 |
| 34 | NLSDGENPKV | site 14 |
| 35 | EISDQRLSI | site 15 |
| 36 | SLGQ | site 16 |
| 37 | TVNPLV | site 17 |

| | | |
|---|---|---|
| 38 | GQSQCPRFNTCPAICAEG | site 18 |
| 39 | DRINWI | site 19 |
| 40 | LDSNATAAN | site 20 |
| 41 | KDNEI | site 21 |
| 42 | QLASEDTNA | site 22 |
| 43 | LKNK | site 23 |
| 44 | DTGDNV | site 24 |
| 45 | KIPEQCT | site 25 |
| 46 | ILSAFNTVIALLGSIVIIVMNIMIIQNYTRSTDNQAVIKDALQGIQQQI<br>KGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSKISQSTASINENVN<br>EKCKFTLPPLKIHECNISCPNPLPFREYRPQTEGVSNLVGLPNNICL<br>QKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMDEGYFAYSHLERI<br>GSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVWTPPNPNTVYHC<br>SAVYNNEFYYVLCAVSTVGDPILNSTYWSGSLMMTRLAVKPKSN<br>GGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQGDTLYFPAVGFLV<br>RTEFKYNDSNCPITKCQYSKPENCRLSMGIRPNSHYILRSGLLKYN<br>LSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVFYQASFSWDTMI<br>KFGDVLTVNPLVVNWRNNTVISRPGQSQCPRFNTCPEICWEGVYN<br>DAFLIDRINWISAGVFLDSNQTAENPVFTVFKDNEILYRAQLASED<br>TNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPKLFAVKIPEQCT | NiV-G backbone ΔCT |
| 47 | ILSAFNTVIALLGSIVIIVMNIMIIQNYTRSTDNQAVIKDALQGIQQQI<br>KGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSKISQSTASINENVN<br>EKCKFTLPPLKIHECNISCPNPLPFREYRPQTEGVSNLVGLPNNICL<br>QKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMDEGYFAYSHLERI<br>GSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVWTPPNPNTVYHC<br>SAVYNNEFYYVLCAVSTVGDPILNSTYWSGSLMMTRLAVKPKSN<br>GGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQGDTLYFPAVGFLV<br>RTEFKYNDSNCPITKCQYSKPENCRLSMGIRPNSHYILRSGLLKYN<br>LSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVFYQASFSWDTMI<br>KFGDVLTVNPLVVNWRNNTVISRPGQSQCPRFNTCPAICAEGVYN<br>DAFLIDRINWISAGVFLDSNATAANPVFTVFKDNEILYRAQLASED<br>TNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPKLFAVKIPEQCT | NiV-G backbone ΔCT and mutated (E501 A, W504A, Q530A, E533A) |
| 48 | MPAENKKVRFENTTSDKGKIPSKVIKSYYGTMDIKKINEGLLDSK | NivG CT_45_FL |
| 49 | MPAENKKVRFENTTSDKGKIPSKVIKSYYGTMDIKKINEGLLDSKI<br>LSAFNTVIALLGSIVIIVMNIMIIQNYTRSTDNQAVIKDALQGIQQQI<br>KGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSKISQSTASINENVN<br>EKCKFTLPPLKIHECNISCPNPLPFREYRPQTEGVSNLVGLPNNICL<br>QKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMDEGYFAYSHLERI<br>GSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVWTPPNPNTVYHC<br>SAVYNNEFYYVLCAVSTVGDPILNSTYWSGSLMMTRLAVKPKSN<br>GGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQGDTLYFPAVGFLV<br>RTEFKYNDSNCPITKCQYSKPENCRLSMGIRPNSHYILRSGLLKYN<br>LSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVFYQASFSWDTMI<br>KFGDVLTVNPLVVNWRNNTVISRPGQSQCPRFNTCPAICAEGVYN<br>DAFLIDRINWISAGVFLDSNATAANPVFTVFKDNEILYRAQLASED<br>TNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPKLFAVKIPEQCT | NiVG with NiVG CT_45_FL, mutated (E501 A, W504A, Q530A, E533A) |
| 50 | MLDSK | NivG_CT_05 cytoplasmic tail |
| 51 | MLLDSK | NivG_CT_06 cytoplasmic tail |
| 52 | MGLLDSK | NivG_CT_07 cytoplasmic tail |

| | | |
|---|---|---|
| 53 | MEGLLDSK | NivG_CT_08 cytoplasmic tail |
| 54 | MNEGLLDSK | NivG_CT_09 cytoplasmic tail |
| 55 | MINEGLLDSK | NivG_CT_10 cytoplasmic tail |
| 56 | MKINEGLLDSK | NivG_CT_11 cytoplasmic tail |
| 57 | MIKKINEGLLDSK | NivG_CT_13 cytoplasmic tail |
| 58 | MDIKKINEGLLDSK | NivG_CT_14 cytoplasmic tail |
| 59 | MMDIKKINEGLLDSK | NivG_CT_15 cytoplasmic tail |
| 60 | MTMDIKKINEGLLDSK | NivG_CT_16 cytoplasmic tail |
| 61 | MGTMDIKKINEGLLDSK | NivG_CT_17 cytoplasmic tail |
| 62 | MYGTMDIKKINEGLLDSK | NivG_CT_18 cytoplasmic tail |
| 63 | MYYGTMDIKKINEGLLDSK | NivG_CT_19 cytoplasmic tail |
| 64 | MSYYGTMDIKKINEGLLDSK | NivG_CT_20 cytoplasmic tail |
| 65 | MKSYYGTMDIKKINEGLLDSK | NivG_CT_21 cytoplasmic tail |
| 66 | MIKSYYGTMDIKKINEGLLDSK | NivG_CT_22 cytoplasmic tail |
| 67 | MVIKSYYGTMDIKKINEGLLDSK | NivG_CT_23 cytoplasmic tail |
| 68 | MKVIKSYYGTMDIKKINEGLLDSK | NivG_CT_24 cytoplasmic tail |
| 69 | MSKVIKSYYGTMDIKKINEGLLDSK | NivG_CT_25 cytoplasmic tail |
| 70 | MKGKNPSKVIKSYYGTMDIKKINEGLLDSK | NivG_CT_30 cytoplasmic tail |
| 71 | MNTTSDKGKNPSKVIKSYYGTMDIKKINEGLLDSK | NivG_CT_35 cytoplasmic tail |
| 72 | MKVRFENTTSDKGKNPSKVIKSYYGTMDIKKINEGLLDSK | NivG_CT_40 cytoplasmic tail |
| 73 | MIIKSYYGTMDIKKINEGLLDSK | NivG_CSUR38_CT_23 cytoplasmic tail |
| 74 | MGKIPSKIIKSYYGTMDIKKINEGLLDSK | NivG_CSUR38_CT_29 cytoplasmic tail |
| 75 | MNTTSDKGKIPSKIIKSYYGTMDIKKINEGLLDSK | NivG_CSUR38_CT_35 cytoplasmic tail |
| 76 | MSKKVRFENTTSDKGKIPSKIIKSYYGTMDIKKINEGLLDSK | NivG_CSUR38_CT_42 cytoplasmic tail |

| | | |
|---|---|---|
| 77 | MPAESKKVRFENTTSDKGKIPSKIIKSYYGTMDIKKINEGLLDSK | NivG_CSUR38_CT_45_FL cytoplasmic tail |
| 78 | M**N**PSKVIKSYYGTMDIKKINEGLLDSK | NivG_VRI-0626_CT_27 cytoplasmic tail |
| 79 | MDKGK**N**PSKVIKSYYGTMDIKKINEGLLDSK | NivG_VRI-0626_CT_31 cytoplasmic tail |
| 80 | MNTTSDKGK**N**PSKVIKSYYGTMDIKKINEGLLDSK | NivG_VRI-0626_CT_35 |
| 81 | MKVRFENTTSDKGK**N**PSKVIKSYYGTMDIKKINEGLLDSK | NivG_VRI-0626_CT_40 cytoplasmic tail |
| 82 | MPAENKKVRFENTTSDKGK**N**PSKVIKSYYGTMDIKKINEGLLDSK | NivG_VRI-0626_CT_45_FL cytoplasmic tail |
| 83 | MKKINEGLLDSK | NivG_CT_12 cytoplasmic tail |
| 84 | MDGLLDSK | HevG_CT_08 |
| 85 | MNDGLLDSK | HevG_CT_09 |
| 86 | MINDGLLDSK | HevG_CT_10 |
| 87 | MKINDGLLDSK | HevG_CT_11 |
| 88 | MKKINDGLLDSK | HevG_CT_12 |
| 89 | MIKKINDGLLDSK | HevG_CT_13 |
| 90 | MDIKKINDGLLDSK | HevG_CT_14 |
| 91 | MMDIKKINDGLLDSK | HevG_CT_15 |
| 92 | MTMDIKKINDGLLDSK | HevG_CT_16 |
| 93 | MGTMDIKKINDGLLDSK | HevG_CT_17 |
| 94 | MYGTMDIKKINDGLLDSK | HevG_CT_18 |
| 95 | MYYGTMDIKKINDGLLDSK | HevG_CT_19 |
| 96 | MNYYGTMDIKKINDGLLDSK | HevG_CT_20 |
| 97 | MKNYYGTMDIKKINDGLLDSK | HevG_CT_21 |
| 98 | MIKNYYGTMDIKKINDGLLDSK | HevG_CT_22 |
| 99 | MVIKNYYGTMDIKKINDGLLDSK | HevG_CT_23 |
| 100 | MKVIKNYYGTMDIKKINDGLLDSK | HevG_CT_24 |
| 101 | MMADSKLVSLNNNLSGKIKDQGKVIKNYYGTMDIKKINDGLLDSK | HevG_CT_45_FL |
| 102 | MQNDHY | BatPV_G_CT_6 |
| 103 | MNQNDHY | BatPV_G_CT_7 |
| 104 | MKNQNDHY | BatPV_G_CT_8 |
| 105 | MQKNQNDHY | BatPV_G_CT_9 |
| 106 | MKQKNQNDHY | BatPV_G_CT_10 |
| 107 | MKKQKNQNDHY | BatPV_G_CT_11 |
| 108 | MWKKQKNQNDHY | BatPV_G_CT_12 |

| | | |
|---|---|---|
| 109 | MNWKKQKNQNDHY | BatPV_G_CT_13 |
| 110 | MRNWKKQKNQNDHY | BatPV_G_CT_14 |
| 111 | MERNWKKQKNQNDHY | BatPV_G_CT_15 |
| 112 | MSERNWKKQKNQNDHY | BatPV_G_CT_16 |
| 113 | MHSERNWKKQKNQNDHY | BatPV_G_CT_17 |
| 114 | MSHSERNWKKQKNQNDHY | BatPV_G_CT_18 |
| 115 | MGSHSERNWKKQKNQNDHY | BatPV_G_CT_19 |
| 116 | MLGSHSERNWKKQKNQNDHY | BatPV_G_CT_20 |
| 117 | MGLGSHSERNWKKQKNQNDHY | BatPV_G_CT_21 |
| 118 | MFGLGSHSERNWKKQKNQNDHY | BatPV_G_CT_22 |
| 119 | MYFGLGSHSERNWKKQKNQNDHY | BatPV_G_CT_23 |
| 120 | MGYFGLGSHSERNWKKQKNQNDHY | BatPV_G_CT_24 |
| 121 | MKITKQGYFGLGSHSERNWKKQKNQNDHY | BatPV_G_CT_29 |
| 122 | MPQKTVEFINMNSPLERGVSTLSDKKTLNQSKITKQGYFGLGSHSERNWKKQKNQNDHY | BatPV_G_CT_59_FL |
| 123 | MLHDIK | CedarV_G_CT_6 |
| 124 | MSLHDIK | CedarV_G_CT_7 |
| 125 | MESLHDIK | CedarV_G_CT_8 |
| 126 | MNESLHDIK | CedarV_G_CT_9 |
| 127 | MLNESLHDIK | CedarV_G_CT_10 |
| 128 | MLLNESLHDIK | CedarV_G_CT_11 |
| 129 | MNLLNESLHDIK | CedarV_G_CT_12 |
| 130 | MSNLLNESLHDIK | CedarV_G_CT_13 |
| 131 | MVSNLLNESLHDIK | CedarV_G_CT_14 |
| 132 | MNVSNLLNESLHDIK | CedarV_G_CT_15 |
| 133 | MYNVSNLLNESLHDIK | CedarV_G_CT_16 |
| 134 | MNYNVSNLLNESLHDIK | CedarV_G_CT_17 |
| 135 | MKNYNVSNLLNESLHDIK | CedarV_G_CT_18 |
| 136 | MNKNYNVSNLLNESLHDIK | CedarV_G_CT_19 |

| | | |
|---|---|---|
| 137 | MKNKNYNVSNLLNESLHDIK | CedarV_G_CT_20 |
| 138 | MVKNKNYNVSNLLNESLHDIK | CedarV_G_CT_21 |
| 139 | MYVKNKNYNVSNLLNESLHDIK | CedarV_G_CT_22 |
| 140 | MYYVKNKNYNVSNLLNESLHDIK | CedarV_G_CT_23 |
| 141 | MSYYVKNKNYNVSNLLNESLHDIK | CedarV_G_CT_24 |
| 142 | MQKDLNKSYYVKNKNYNVSNLLNESLHDIK | CedarV_G_CT_30 |
| 143 | MLSQLQKNYLDNSNQQGDKMNNPDKKLSVNFNPLELDKGQKDLNKSYYVKNKNYNVSNLLNESLHDIK | CedarV_G_CT_68_FL |
| 144 | MQGGRR | CDV-H_CT_06 |
| 145 | MEQGGRR | CDV-H_CT_07 |
| 146 | MEEQGGRR | CDV-H_CT_08 |
| 147 | MTEEQGGRR | CDV-H_CT_09 |
| 148 | MVTEEQGGRR | CDV-H_CT_10 |
| 149 | MLVTEEQGGRR | CDV-H_CT_11 |
| 150 | MSLVTEEQGGRR | CDV-H_CT_12 |
| 151 | MLSLVTEEQGGRR | CDV-H_CT_13 |
| 152 | MKLSLVTEEQGGRR | CDV-H_CT_14 |
| 153 | MSKLSLVTEEQGGRR | CDV-H_CT_15 |
| 154 | MSSKLSLVTEEQGGRR | CDV-H_CT_16 |
| 155 | MNSSKLSLVTEEQGGRR | CDV-H_CT_17 |
| 156 | MANSSKLSLVTEEQGGRR | CDV-H_CT_18 |
| 157 | MRANSSKLSLVTEEQGGRR | CDV-H_CT_19 |
| 158 | MARANSSKLSLVTEEQGGRR | CDV-H_CT_20 |
| 159 | MLSYQDKVGAFYKDNARANSSKLSLVTEEQGGRR | CDV-H_CT_34_FL |
| 160 | MRIIFR | HPIV2-HN_CT_06 |
| 161 | MIPKRTCRIIFR | HPIV2-HN_CT_12 |
| 162 | MEDYSNLSLKSIPKRTCRIIFR | HPIV2-HN_CT_22_FL |
| 163 | MLMIDR | MvH_CT_06 |
| 164 | MHLMIDR | MvH_CT_07 |
| 165 | MEHLMIDR | MvH_CT_08 |
| 166 | MREHLMIDR | MvH_CT_09 |
| 167 | MNREHLMIDR | MvH_CT_10 |
| 168 | MINREHLMIDR | MvH_CT_11 |
| 169 | MVINREHLMIDR | MvH_CT_12 |
| 170 | MIVINREHLMIDR | MvH_CT_13 |
| 171 | MRIVINREHLMIDR | MvH_CT_14 |
| 172 | MSRIVINREHLMIDR | MvH_CT_15 |
| 173 | MGSRIVINREHLMIDR | MvH_CT_16 |

| | | |
|---|---|---|
| 174 | MKGSRIVINREHLMIDR | MvH_CT_17 |
| 175 | MPKGSRIVINREHLMIDR | MvH_CT_18 |
| 176 | MHPKGSRIVINREHLMIDR | MvH_CT_19 |
| 177 | MPHPKGSRIVINREHLMIDR | MvH_CT_20 |
| 178 | MSPQRDRINAFYKDNPHPKGSRIVINREHLMIDR | MvH_CT_34_FL |
| 179 | MRLVFR | NDV-HN_CT_06 |
| 180 | MWRLVFR | NDV-HN_CT_07 |
| 181 | MTWRLVFR | NDV-HN_CT_08 |
| 182 | MNTWRLVFR | NDV-HN_CT_09 |
| 183 | MKNTWRLVFR | NDV-HN_CT_10 |
| 184 | MAKNTWRLVFR | NDV-HN_CT_11 |
| 185 | MEAKNTWRLVFR | NDV-HN_CT_12 |
| 186 | MREAKNTWRLVFR | NDV-HN_CT_13 |
| 187 | MEREAKNTWRLVFR | NDV-HN_CT_14 |
| 188 | MDEREAKNTWRLVFR | NDV-HN_CT_15 |
| 189 | MNDEREAKNTWRLVFR | NDV-HN_CT_16 |
| 190 | MENDEREAKNTWRLVFR | NDV-HN_CT_17 |
| 191 | MLENDEREAKNTWRLVFR | NDV-HN_CT_18 |
| 192 | MALENDEREAKNTWRLVFR | NDV-HN_CT_19 |
| 193 | MVALENDEREAKNTWRLVFR | NDV-HN_CT_20 |
| 194 | MERGVSQVALENDEREAKNTWRLVFR | NDV-HN_CT_26_FL |
| 195 | MERSGK | SeV_CT_06 |
| 196 | MSERSGK | SeV_CT_07 |
| 197 | MDSERSGK | SeV_CT_08 |
| 198 | MSDSERSGK | SeV_CT_09 |
| 199 | MVSDSERSGK | SeV_CT_10 |
| 200 | MLVSDSERSGK | SeV_CT_11 |
| 201 | MKLVSDSERSGK | SeV_CT_12 |
| 202 | MTKLVSDSERSGK | SeV_CT_13 |
| 203 | MTTKLVSDSERSGK | SeV_CT_14 |
| 204 | MSTTKLVSDSERSGK | SeV_CT_15 |
| 205 | MGSTTKLVSDSERSGK | SeV_CT_16 |
| 206 | MGGSTTKLVSDSERSGK | SeV_CT_17 |
| 207 | MPGGSTTKLVSDSERSGK | SeV_CT_18 |

| | | |
|---|---|---|
| 208 | MSPGGSTTKLVSDSERSGK | SeV_CT_19 |
| 209 | MTSPGGSTTKLVSDSERSGK | SeV_CT_20 |
| 210 | MDGDRSKRDSYWSTSPGGSTTKLVSDSERSGK | SeV_CT_32_FL |
| 211 | MDQAEEDTRLVQYQQTLVMAHIINLKDNIFATLRN | BaEVwt_CT |
| 212 | MAHIINLKDNIFATLRNKKINEGLLDSK | BaEVRLess_CT |
| 213 | MAHIINLKDNIFATLRN | BaEVRLess_CT |
| 214 | MKRFRSMEIDNYIRKNNSGQYRYR | Cocal_CT |
| 215 | MKRFRSMEIDNYIRKNNSGQYRYRKKINEGLLDSK | Cocal_CT |
| 216 | MFVFKC | EboV-GP_CT |
| 217 | MFVFKCKKINEGLLDSK | EboV-GP_CT |
| 218 | MLNGENELAQYKQRLVLIKVASIRDNIFQVLKKKINEGLLDSK | GaLV_CT |
| 219 | MLNGENELAQYKQRLVLIKVASIRDNIFQVLK | GaLV_CT |
| 220 | MLIKVASIRDNIFQVLKKKINEGLLDSK | GaLV-RLess_CT |
| 221 | MLIKVASIRDNIFQVLK | GaLV-RLess_CT |
| 222 | MRRKWVTKVGPVKFAGCSCIGKNTLRHPKPCSGGKIHRHT | LCMV_GP_CT |
| 223 | MPEYEIPKLQHYQQTLVLAQVVSIRDKVFQVLRKKINEGLLDSK | MLV-A_CT |
| 224 | MPEYEIPKLQHYQQTLVLAQVVSIRDKVFQVLR | MLV-A_CT |
| 225 | MLAQVVSIRDKVFQVLRKKINEGLLDSK | MLV-A-RLess_CT |
| 226 | MLAQVVSIRDKVFQVLR | MLV-A-RLess_CT |
| 227 | MKGLRNMEIDTYIQRKKTHKLKICLHIGVR | VSV-G_CT |
| 228 | MKGLRNMEIDTYIQRKKTHKLKICLHIGVRKKINEGLLDSK | VSV-G_CT |
| 229 | MLLIRELGQRIRRPIHRIARYAAQLVEIVRDTGEAVAIATANLLNV ASNKLEQSWYQLLNWWYKLAEWGRRGLLEVIRTVILLLDRLRHY SFLCLSRLDDWILALSGNVLRISRDRDREGGEEEIGEPRDPGRPIPL HTQFSLPSYGQRVRN | HIV-1_gp41_MLLIR-LLP1-LLP3-LLP2-gp41-CT-N |
| 230 | MLLIRELGQRIRRPIHRIARYAAQLVEIVRDTGEAVAIATANLLNV ASNKLEQSWYQLLNWWYKLAEWGRRGLLEVIRTVILLLDRLRHY SFLCLSRLDDWILALSGNVLRISRDRDREGGEEEIGEPRDPGRPIPL HTQFSLPSYGQRVRNKKINEGLLDSK | HIV-1_gp41_MLLIR-LLP1-LLP3-LLP2-gp41-CT-N |
| 231 | MAVAIATANLLNVASNKLEQSWYQLLNWWYKLAEWGRRGLLEV IRTVILLLDRLRHYSFLCLSRLDDWILALSGNVLRISRDRDREGGEE EIGEPRDPGRPIPLHTQFSLPSYGQRVRN | HIV-1_gp41_LLP3-LLP2-gp41-CT-N |
| 232 | MLLEVIRTVILLLDRLRHYSFLCLSRLDDWILALSGNVLRISRDRDR EGGEEEIGEPRDPGRPIPLHTQFSLPSYGQRVRN | HIV-1_gp41_LLP2-gp41-CT-N |
| 233 | MSGNVLRISRDRDREGGEEEIGEPRDPGRPIPLHTQFSLPSYGQRVR N | HIV-1_gp41_gp41-CT-N |
| 234 | MLLIRELGQRIRRPIHRIARYAAQLVEIVRDTGEAVAIATANLLNV ASNKLEQSWYQLLNWWYKLAEWGRRGLLEVIRTVILLLDRLRHY SFLCLSRLDDWILAL | HIV-1_gp41_MLLIR-LLP1-LLP3-LLP2 |
| 235 | MELGQRIRRPIHRIARYAAQLVEIVRDTGEAVAIATANLLNVASNK LEQSWYQLLNWWYKLAEWGRRGLLEVIRTVILLLDRLRHYSFLC LSRLDDWILAL | HIV-1_gp41_LLP1-LLP3-LLP2 |

| | | |
|---|---|---|
| 236 | MAVAIATANLLNVASNKLEQSWYQLLNWWYKLAEWGRRGLLEV IRTVILLLDRLRHYSFLCLSRLDDWILAL | HIV-1_gp41_LLP3-LLP2 |
| 237 | MLLEVIRTVILLLDRLRHYSFLCLSRLDDWILAL | HIV-1_gp41_LLP2 |
| 238 | MLLIRELGQRIRRPIHRIARYAAQLVEIVRDTGEAVAIATANLLNV ASNKLEQSWYQLLNWWYKLAEW | HIV-1_gp41_MLLIR-LLP1-LLP3 |
| 239 | MELGQRIRRPIHRIARYAAQLVEIVRDTGEAVAIATANLLNVASNK LEQSWYQLLNWWYKLAEW | HIV-1_gp41_LLP1-LLP3 |
| 240 | MAVAIATANLLNVASNKLEQSWYQLLNWWYKLAEW | HIV-1_gp41_LLP3 |
| 241 | MLLIRELGQRIRRPIHRIARYAAQLVEIVR | HIV-1_gp41_MLLIR-LLP1 |
| 242 | MELGQRIRRPIHRIARYAAQLVEIVR | HIV-1_gp41_LLP1 |
| 243 | MLLIR | HIV-1_gp41_MLLIR |
| 244 | MAVEGGMKCVK | CD63_CT_1to11 |
| 245 | MCGACKENYC | CD63_CT_73to81 |
| 246 | MVEYGSRISKVLCC | CD63_CT_225to238 |
| 247 | MAVEGGMKCVKKKINEGLLDSK | CD63_CT_1to11 |
| 248 | MCGACKENYCKKINEGLLDSK | CD63_CT_73to81 |
| 249 | MVEYGSRISKVLCCKKINEGLLDSK | CD63_CT_225to238 |
| 250 | MKGEYKPNVVTTVASKYIPNEGTDWKANMKEKEFKAFERRDHII MLLK | CD29_CT |
| 251 | MKGEYKPNVVTTVASKYIPNEGTDWKANMKEKEFKAFERRDHII MLLKKKINEGLLDSK | CD29_CT |
| 252 | MVINREHLMIDRILSAFNTVIALLGSIVIIVMNIMIIQNYTRSTDNQA VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK ISQSTASINENVNEKCKFTLPPLKIHECNISCPNPLPFREYRPQTEGV SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILNSTYWSGSLM MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG DTLYFPAVGFLVRTEFKYNDSNCPITKCQYSKPENCRLSMGIRPNS HYILRSGLLKYNLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF YQASFSWDTMIKFGDVLTVNPLVVNWRNNTVISRPGQSQCPRFNT CPAICAEGVYNDAFLIDRINWISAGVFLDSNATAANPVFTVFKDNE ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK LFAVKIPEQCT | MvH_CT_12 |
| 253 | MGSRIVINREHLMIDRILSAFNTVIALLGSIVIIVMNIMIIQNYTRSTD NQAVIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLL GSKISQSTASINENVNEKCKFTLPPLKIHECNISCPNPLPFREYRPQT EGVSNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLL AMDEGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMT NVWTPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILNSTYWSG | MvH_CT_16 |

| | | |
|---|---|---|
| | SLMMTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGI KQGDTLYFPAVGFLVRTEFKYNDSNCPITKCQYSKPENCRLSMGIR PNSHYILRSGLLKYNLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQ PVFYQASFSWDTMIKFGDVLTVNPLVVNWRNNTVISRPGQSQCPR FNTCPAICAEGVYNDAFLIDRINWISAGVFLDSNATAANPVFTVFK DNEILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVI RPKLFAVKIPEQCT | |
| 254 | MPHPKGSRIVINREHLMIDRILSAFNTVIALLGSIVIIVMNIMIIQNYT RSTDNQAVIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPA NIGLLGSKISQSTASINENVNEKCKFTLPPLKIHECNISCPNPLPFRE YRPQTEGVSNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCI TDPLLAMDEGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVP SLFMTNVWTPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILNST YWSGSLMMTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPY GPSGIKQGDTLYFPAVGFLVRTEFKYNDSNCPITKCQYSKPENCRL SMGIRPNSHYILRSGLLKYNLSDGENPKVVFIEISDQRLSIGSPSKIY DSLGQPVFYQASFSWDTMIKFGDVLTVNPLVVNWRNNTVISRPGQ SQCPRFNTCPAICAEGVYNDAFLIDRINWISAGVFLDSNATAANPV FTVFKDNEILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDT GDNVIRPKLFAVKIPEQCT | MvH_CT_20 |
| 255 | MYYGTMDIKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQNYT RSTDNQAVIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPA NIGLLGSKISQSTASINENVNEKCKFTLPPLKIHECNISCPNPLPFRE YRPQTEGVSNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCI TDPLLAMDEGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVP SLFMTNVWTPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILNST YWSGSLMMTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPY GPSGIKQGDTLYFPAVGFLVRTEFKYNDSNCPITKCQYSKPENCRL SMGIRPNSHYILRSGLLKYNLSDGENPKVVFIEISDQRLSIGSPSKIY DSLGQPVFYQASFSWDTMIKFGDVLTVNPLVVNWRNNTVISRPGQ SQCPRFNTCPAICAEGVYNDAFLIDRINWISAGVFLDSNATAANPV FTVFKDNEILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDT GDNVIRPKLFAVKIPEQCT | NivG_CT_19 |
| 256 | MKNQNDHYILSAFNTVIALLGSIVIIVMNIMIIQNYTRSTDNQAVIK DALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSKISQ STASINENVNEKCKFTLPPLKIHECNISCPNPLPFREYRPQTEGVSNL VGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMDEGY FAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVWTPP NPNTVYHCSAVYNNEFYYVLCAVSTVGDPILNSTYWSGSLMMTR LAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQGDTL YFPAVGFLVRTEFKYNDSNCPITKCQYSKPENCRLSMGIRPNSHYI LRSGLLKYNLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVFYQ ASFSWDTMIKFGDVLTVNPLVVNWRNNTVISRPGQSQCPRFNTCP AICAEGVYNDAFLIDRINWISAGVFLDSNATAANPVFTVFKDNEIL YRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPKLF AVKIPEQCT | BatPV_G_CT_8 |
| 257 | MDSERSGKILSAFNTVIALLGSIVIIVMNIMIIQNYTRSTDNQAVIKD ALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSKISQS TASINENVNEKCKFTLPPLKIHECNISCPNPLPFREYRPQTEGVSNL VGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMDEGY FAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVWTPP NPNTVYHCSAVYNNEFYYVLCAVSTVGDPILNSTYWSGSLMMTR LAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQGDTL | SeV_CT_08 |

| | | |
|---|---|---|
| | YFPAVGFLVRTEFKYNDSNCPITKCQYSKPENCRLSMGIRPNSHYI LRSGLLKYNLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVFYQ ASFSWDTMIKFGDVLTVNPLVVNWRNNTVISRPGQSQCPRFNTCP AICAEGVYNDAFLIDRINWISAGVFLDSNATAANPVFTVFKDNEIL YRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPKLF AVKIPEQCT | |
| 258 | MLAQVVSIRDKVFQVLRILSAFNTVIALLGSIVIIVMNIMIIQNYTRS TDNQAVIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANI GLLGSKISQSTASINENVNEKCKFTLPPLKIHECNISCPNPLPFREYR PQTEGVSNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITD PLLAMDEGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSL FMTNVWTPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILNSTY WSGSLMMTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYG PSGIKQGDTLYFPAVGFLVRTEFKYNDSNCPITKCQYSKPENCRLS MGIRPNSHYILRSGLLKYNLSDGENPKVVFIEISDQRLSIGSPSKIYD SLGQPVFYQASFSWDTMIKFGDVLTVNPLVVNWRNNTVISRPGQS QCPRFNTCPAICAEGVYNDAFLIDRINWISAGVFLDSNATAANPVF TVFKDNEILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTG DNVIRPKLFAVKIPEQCT | MLV-A-RLess_CT |
| 259 | MLLIRILSAFNTVIALLGSIVIIVMNIMIIQNYTRSTDNQAVIKDALQ GIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSKISQSTASI NENVNEKCKFTLPPLKIHECNISCPNPLPFREYRPQTEGVSNLVGLP NNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMDEGYFAYS HLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVWTPPNPNT VYHCSAVYNNEFYYVLCAVSTVGDPILNSTYWSGSLMMTRLAVK PKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQGDTLYFPAV GFLVRTEFKYNDSNCPITKCQYSKPENCRLSMGIRPNSHYILRSGLL KYNLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVFYQASFSWD TMIKFGDVLTVNPLVVNWRNNTVISRPGQSQCPRFNTCPAICAEG VYNDAFLIDRINWISAGVFLDSNATAANPVFTVFKDNEILYRAQLA SEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPKLFAVKIPE QCT | HIV-1_gp41_MLLIR |
| 260 | MAVEGGMKCVKILSAFNTVIALLGSIVIIVMNIMIIQNYTRSTDNQA VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK ISQSTASINENVNEKCKFTLPPLKIHECNISCPNPLPFREYRPQTEGV SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILNSTYWSGSLM MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG DTLYFPAVGFLVRTEFKYNDSNCPITKCQYSKPENCRLSMGIRPNS HYILRSGLLKYNLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF YQASFSWDTMIKFGDVLTVNPLVVNWRNNTVISRPGQSQCPRFNT CPAICAEGVYNDAFLIDRINWISAGVFLDSNATAANPVFTVFKDNE ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK LFAVKIPEQCT | CD63_CT_1to11 |
| 261 | MCGACKENYCILSAFNTVIALLGSIVIIVMNIMIIQNYTRSTDNQAV IKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSKIS QSTASINENVNEKCKFTLPPLKIHECNISCPNPLPFREYRPQTEGVS NLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMDE GYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVWT PPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILNSTYWSGSLMM TRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQGD TLYFPAVGFLVRTEFKYNDSNCPITKCQYSKPENCRLSMGIRPNSH | CD63_CT_73to81 |

| | | |
|---|---|---|
| | YILRSGLLKYNLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVFY QASFSWDTMIKFGDVLTVNPLVVNWRNNTVISRPGQSQCPRFNTC PAICAEGVYNDAFLIDRINWISAGVFLDSNATAANPVFTVFKDNEI LYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK LFAVKIPEQCT | |
| 262 | MKKINDGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQNYTRSTDNQA VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK ISQSTASINENVNEKCKFTLPPLKIHECNISCPNPLPFREYRPQTEGV SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILNSTYWSGSLM MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG DTLYFPAVGFLVRTEFKYNDSNCPITKCQYSKPENCRLSMGIRPNS HYILRSGLLKYNLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF YQASFSWDTMIKFGDVLTVNPLVVNWRNNTVISRPGQSQCPRFNT CPAICAEGVYNDAFLIDRINWISAGVFLDSNATAANPVFTVFKDNE ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK LFAVKIPEQCT | HevG_CT_12 |
| 263 | MPEYEIPKLQHYQQTLVLAQVVSIRDKVFQVLRILSAFNTVIALLG SIVIIVMNIMIIQNYTRSTDNQAVIKDALQGIQQQIKGLADKIGTEIG PKVSLIDTSSTITIPANIGLLGSKISQSTASINENVNEKCKFTLPPLKI HECNISCPNPLPFREYRPQTEGVSNLVGLPNNICLQKTSNQILKPKLI SYTLPVVGQSGTCITDPLLAMDEGYFAYSHLERIGSCSRGVSKQRII GVGEVLDRGDEVPSLFMTNVWTPPNPNTVYHCSAVYNNEFYYVL CAVSTVGDPILNSTYWSGSLMMTRLAVKPKSNGGGYNQHQLALR SIEKGRYDKVMPYGPSGIKQGDTLYFPAVGFLVRTEFKYNDSNCPI TKCQYSKPENCRLSMGIRPNSHYILRSGLLKYNLSDGENPKVVFIEI SDQRLSIGSPSKIYDSLGQPVFYQASFSWDTMIKFGDVLTVNPLVV NWRNNTVISRPGQSQCPRFNTCPAICAEGVYNDAFLIDRINWISAG VFLDSNATAANPVFTVFKDNEILYRAQLASEDTNAQKTITNCFLLK NKIWCISLVEIYDTGDNVIRPKLFAVKIPEQCT | MLV-A_CT |
| 264 | MLLDSKILSAFNTVIALLGSIVIIVMNIMIIQNYTRSTDNQAVIKDAL QGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSKISQSTA SINENVNEKCKFTLPPLKIHECNISCPNPLPFREYRPQTEGVSNLVG LPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMDEGYFA YSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVWTPPNP NTVYHCSAVYNNEFYYVLCAVSTVGDPILNSTYWSGSLMMTRLA VKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQGDTLYFP AVGFLVRTEFKYNDSNCPITKCQYSKPENCRLSMGIRPNSHYILRS GLLKYNLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVFYQASFS WDTMIKFGDVLTVNPLVVNWRNNTVISRPGQSQCPRFNTCPAICA EGVYNDAFLIDRINWISAGVFLDSNATAANPVFTVFKDNEILYRAQ LASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPKLFAVKI PEQCT | NivG_CT_06 |
| 265 | MIKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQNYTRSTDNQ AVIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGS KISQSTASINENVNEKCKFTLPPLKIHECNISCPNPLPFREYRPQTEG VSNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAM DEGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNV WTPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILNSTYWSGSL MMTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQ GDTLYFPAVGFLVRTEFKYNDSNCPITKCQYSKPENCRLSMGIRPN SHYILRSGLLKYNLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPV | NivG_CT_13 |

| | | |
|---|---|---|
| | FYQASFSWDTMIKFGDVLTVNPLVVNWRNNTVISRPGQSQCPRFN TCPAICAEGVYNDAFLIDRINWISAGVFLDSNATAANPVFTVFKDN EILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRP KLFAVKIPEQCT | |
| 266 | MWKKQKNQNDHYILSAFNTVIALLGSIVIIVMNIMIIQNYTRSTDN QAVIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLG SKISQSTASINENVNEKCKFTLPPLKIHECNISCPNPLPFREYRPQTE GVSNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLA MDEGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTN VWTPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILNSTYWSGSL MMTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQ GDTLYFPAVGFLVRTEFKYNDSNCPITKCQYSKPENCRLSMGIRPN SHYILRSGLLKYNLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPV FYQASFSWDTMIKFGDVLTVNPLVVNWRNNTVISRPGQSQCPRFN TCPAICAEGVYNDAFLIDRINWISAGVFLDSNATAANPVFTVFKDN EILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRP KLFAVKIPEQCT | BatPV_G_CT_1 2 |
| 267 | MRIIFRILSAFNTVIALLGSIVIIVMNIMIIQNYTRSTDNQAVIKDALQ GIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSKISQSTASI NENVNEKCKFTLPPLKIHECNISCPNPLPFREYRPQTEGVSNLVGLP NNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMDEGYFAYS HLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVWTPPNPNT VYHCSAVYNNEFYYVLCAVSTVGDPILNSTYWSGSLMMTRLAVK PKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQGDTLYFPAV GFLVRTEFKYNDSNCPITKCQYSKPENCRLSMGIRPNSHYILRSGLL KYNLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVFYQASFSWD TMIKFGDVLTVNPLVVNWRNNTVISRPGQSQCPRFNTCPAICAEG VYNDAFLIDRINWISAGVFLDSNATAANPVFTVFKDNEILYRAQLA SEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPKLFAVKIPE QCT | HPIV2-HN_CT_06 |
| 268 | MIPKRTCRIIFRILSAFNTVIALLGSIVIIVMNIMIIQNYTRSTDNQAVI KDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSKIS QSTASINENVNEKCKFTLPPLKIHECNISCPNPLPFREYRPQTEGVS NLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMDE GYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVWT PPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILNSTYWSGSLMM TRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQGD TLYFPAVGFLVRTEFKYNDSNCPITKCQYSKPENCRLSMGIRPNSH YILRSGLLKYNLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVFY QASFSWDTMIKFGDVLTVNPLVVNWRNNTVISRPGQSQCPRFNTC PAICAEGVYNDAFLIDRINWISAGVFLDSNATAANPVFTVFKDNEI LYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK LFAVKIPEQCT | HPIV2-HN_CT_12 |
| 269 | MKLVSDSERSGKILSAFNTVIALLGSIVIIVMNIMIIQNYTRSTDNQA VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK ISQSTASINENVNEKCKFTLPPLKIHECNISCPNPLPFREYRPQTEGV SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILNSTYWSGSLM MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG DTLYFPAVGFLVRTEFKYNDSNCPITKCQYSKPENCRLSMGIRPNS HYILRSGLLKYNLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF YQASFSWDTMIKFGDVLTVNPLVVNWRNNTVISRPGQSQCPRFNT | CD63_CT_1to11 |

| | | |
|---|---|---|
| | CPAICAEGVYNDAFLIDRINWISAGVFLDSNATAANPVFTVFKDNE ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK LFAVKIPEQCT | |
| 270 | MMADSKLVSLNNNLSGKIKDQGKVIKNYYGTMDIKKINDGLLDS KILGAFNTVIALLGSIIIIVMNIMIIQNYTRTTDNQALIKESLQSVQQ QIKALTDKIGTEIGPKVSLIDTSSTITIPANIGLLGSKISQSTSSINENV NDKCKFTLPPLKIHECNISCPNPLPFREYRPISQGVSDLVGLPNQICL QKTTSTILKPRLISYTLPINTREGVCITDPLLAVDNGFFAYSHLEKIG SCTRGIAKQRIIGVGEVLDRGDKVPSMFMTNVWTPPNPSTIHHCSS TYHEDFYYTLCAVSHVGDPILNSTSWTESLSLIRLAVRPKSDSGDY NQKYIAITKVERGKYDKVMPYGPSGIKQGDTLYFPAVGFLPRTEF QYNDSNCPIIHCKYSKAENCRLSMGVNSKSHYILRSGLLKYNLSLG GDIILQFIEIADNRLTIGSPSKIYNSLGQPVFYQASYSWDTMIKLGD VDTVDPLRVQWRNNSVISRPGQSQCPRFNVCPEVCWEGTYNDAF LIDRLNWVSAGVYLNSNQTAENPVFAVFKDNEILYQVPLAEDDTN AQKTITDCFLLENVIWCISLVEIYDTGDSVIRPKLFAVKIPAQCSES | gb:AF017149\|Organism:Hendra virus\|Strain Name:UNKNOWN-AF017149\|Protein Name:glycoprotein\|Gene Symbol:G |
| 271 | MLSQLQKNYLDNSNQQGDKMNNPDKKLSVNFNPLELDKGQKDL NKSYYVKNKNYNVSNLLNESLHDIKFCIYCIFSLLIIITIINIITISIVIT RLKVHEENNGMESPNLQSIQDSLSSLTNMINTEITPRIGILVTATSVT LSSSINYVGTKTNQLVNELKDYITKSCGFKVPELKLHECNISCADP KISKSAMYSTNAYAELAGPPKIFCKSVSKDPDFRLKQIDYVIPVQQ DRSICMNNPLLDISDGFFTYIHYEGINSCKKSDSFKVLLSHGEIVDR GDYRPSLYLLSSHYHPYSMQVINCVPVTCNQSSFVFCHISNNTKTL DNSDYSSDEYYITYFNGIDRPKTKKIPINNMTADNRYIHFTFSGGG GVCLGEEFIIPVTTVINTDVFTHDYCESFNCSVQTGKSLKEICSESLR SPTNSSRYNLNGIMIISQNNMTDFKIQLNGITYNKLSFGSPGRLSKT LGQVLYYQSSMSWDTYLKAGFVEKWKPFTPNWMNNTVISRPNQ GNCPRYHKCPEICYGGTYNDIAPLDLGKDMYVSVILDSDQLAENP EITVFNSTTILYKERVSKDELNTRSTTTSCFLFLDEPWCISVLETNRF NGKSIRPEIYSYKIPKYC | gb:JQ001776:8170-10275\|Organism:Cedar virus\|Strain Name:CG1a\|Protein Name:attachment glycoprotein\|Gene Symbol:G |
| 272 | MPQKTVEFINMNSPLERGVSTLSDKKTLNQSKITKQGYFGLGSHSE RNWKKQKNQNDHYMTVSTMILEILVVLGIMFNLIVLTMVYYQND NINQRMAELTSNITVLNLNLNQLTNKIQREIIPRITLIDTATTITIPSAI TYILATLTTRISELLPSINQKCEFKTPTLVLNDCRINCTPPLNPSDGV KMSSLATNLVAHGPSPCRNFSSVPTIYYYRIPGLYNRTALDERCILN PRLTISSTKFAYVHSEYDKNCTRGFKYYELMTFGEILEGPEKEPRM FSRSFYSPTNAVNYHSCTPIVTVNEGYFLCLECTSSDPLYKANLSNS TFHLVILRHNKDEKIVSMPSFNLSTDQEYVQIIPAEGGGTAESGNLY FPCIGRLLHKRVTHPLCKKSNCSRTDDESCLKSYYNQGSPQHQVV NCLIRIRNAQRDNPTWDVITVDLTNTYPGSRSRIFGSFSKPMLYQSS VSWHTLLQVAEITDLDKYQLDWLDTPYISRPGGSECPFGNYCPTV CWEGTYNDVYSLTPNNDLFVTVYLKSEQVAENPYFAIFSRDQILK EFPLDAWISSARTTTISCFMFNNEIWCIAALEITRLNDDIIRPIYYSF WLPTDCRTPYPHTGKMTRVPLRSTYNY | gb:NC_025256:9117-11015\|Organism:Bat Paramyxovirus Eid_hel/GH-M74a/GHA/2009\|Strain Name:BatPV/Eid_hel/GH-M74a/GHA/2009\|Protein Name:glycoprotein\|Gene Symbol:G |
| 273 | MATNRDNTITSAEVSQEDKVKKYYGVETAEKVADSISGNKVFILM NTLLILTGAIITITLNITNLTAAKSQQNMLKIIQDDVNAKLEMFVNL DQLVKGEIKPKVSLINTAVSVSIPGQISNLQTKFLQKYVYLEESITK QCTCNPLSGIFPTSGPTYPPTDKPDDDTTDDDKVDTTIKPIEYPKPD GCNRTGDHFTMEPGANFYTVPNLGPASSNSDECYTNPSFSIGSSIY MFSQEIRKTDCTAGEILSIQIVLGRIVDKGQQGPQASPLLVWAVPN PKIINSCAVAAGDEMGWVLCSVTLTAASGEPIPHMFDGFWLYKLE PDTEVVSYRITGYAYLLDKQYDSVFIGKGGGIQKGNDLYFQMYGL | gb:NC_025352:8716-11257\|Organism:Mojiang virus\|Strain Name:Tongguan1\|Protein Name:attachment |

| | | |
|---|---|---|
| | SRNRQSFKALCEHGSCLGTGGGGYQVLCDRAVMSFGSEESLITNA YLKVNDLASGKPVIIGQTFPPSDSYKGSNGRMYTIGDKYGLYLAPS SWNRYLRFGITPDISVRSTTWLKSQDPIMKILSTCTNTDRDMCPEIC NTRGYQDIFPLSEDSEYYTYIGITPNNGGTKNFVAVRDSDGHIASID ILQNYYSITSATISCFMYKDEIWCIAITEGKKQKDNPQRIYAHSYKI RQMCYNMKSATVTVGNAKNITIRRY | glycoprotein\|Gene Symbol:G |
| 274 | DIQMTQTTSSLSASLGDRVTISCRASQDISKYLNWYQQKPDGTVKL LIYHTSRLHSGVPSRFSGSGSGTDYSLTISNLEQEDIATYFCQQGNT LPYTFGGGTKLEITGSTSGSGKPGSGEGSTKGEVKLQESGPGLVAP SQSLSVTCTVSGVSLPDYGVSWIRQPPRKGLEWLGVIWGSETTYY NSALKSRLTIIKDNSKSQVFLKMNSLQTDDTAIYYCAKHYYYGGS YAMDYWGQGTSVTVSS | Anti-CD19 scFv (FMC63) |
| 275 | DIQMTQTTSSLSASLGDRVTISCRASQDISKYLNWYQQKPDGTVKL LIYHTSRLHSGVPSRFSGSGSGTDYSLTISNLEQEDIATYFCQQGNT LPYTFGGGTKLEITGGGGSGGGGSGGGGSEVKLQESGPGLVAPSQ SLSVTCTVSGVSLPDYGVSWIRQPPRKGLEWLGVIWGSETTYYNS ALKSRLTIIKDNSKSQVFLKMNSLQTDDTAIYYCAKHYYYGGSYA MDYWGQGTSVTVSS | Anti-CD19 scFv (FMC63) |
| 276 | ESKYGPPCPPCP | IgG4 Hinge |
| 277 | TTTPAPRPPTPAPTIASQPLSLRPE | CD8 Hinge |
| 278 | IEVMYPPPYLDNEKSNGTIIHVKGKHLCPSPLFPGPSKP | CD28 |
| 279 | ACRPAAGGAVHTRGLDFACDIYIWAPLAGTCGVLLLSLVITLYC | CD8 |
| 280 | FWVLVVVGGVLACYSLLVTVAFIIFWV | CD28 |
| 281 | FWVLVVVGGVLACYSLLVTVAFIIFWV | CD28 |
| 282 | RSKRSRLLHSDYMNMTPRRPGPTRKHYQPYAPPRDFAAYRS | CD28 |
| 283 | KRGRKKLLYIFKQPFMRPVQTTQEEDGCSCRFPEEEEGGCEL | 4-1BB |
| 284 | RVKFSRSADAPAYQQGQNQLYNELNLGRREEYDVLDKRRGRDPE MGGKPRRKNPQEGLYNELQKDKMAEAYSEIGMKGERRRGKGHD GLYQGLSTATKDTYDALHMQALPPR | CD3zeta |
| 285 | RVKFSRSADAPAYKQGQNQLYNELNLGRREEYDVLDKRRGRDPE MGGKPRRKNPQEGLYNELQKDKMAEAYSEIGMKGERRRGKGHD GLYQGLSTATKDTYDALHMQALPPR | CD3zeta |
| 286 | MALPVTALLLPLALLLHAARP | CD8α signal peptide |
| 287 | METDTLLLWVLLLWVPGSTG | IgK signal peptide |
| 288 | MLLLVTSLLLCELPHPAFLLIP | GMCSFR-α (CSF2RA) signal peptide |
| 289 | TTTPAPRPPTPAPTIASQPLSLRPEACRPAAGGAVHTRGLDFACD | CD8α hinge domain |
| 290 | IEVMYPPPYLDNEKSNGTIIHVKGKHLCPSPLFPGPSKP | CD28 hinge domain |
| 291 | AAAIEVMYPPPYLDNEKSNGTIIHVKGKHLCPSPLFPGPSKP | CD28 hinge domain |
| 292 | ESKYGPPCPPCP | IgG4 hinge domain |
| 293 | ESKYGPPCPSCP | IgG4 hinge domain |
| 294 | ESKYGPPCPPCPAPEFLGGPSVFLFPPKPKDTLMISRTPEVTCVVVD VSQEDPEVQFNWYVDGVEVHNAKTKPREEQFNSTYRVVSVLTVL | IgG4 hinge-CH2-CH3 domain |

| | | |
|---|---|---|
| | HQDWLNGKEYKCKVSNKGLPSSIEKTISKAKGQPREPQVYTLPPS QEEMTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPVLD SDGSFFLYSRLTVDKSRWQEGNVFSCSVMHEALHNHYTQKSLSLS LGK | |
| 295 | IYIWAPLAGTCGVLLLSLVITLYC | CD8α transmembrane domain |
| 296 | FWVLVVVGGVLACYSLLVTVAFIIFWV | CD28 transmembrane domain |
| 297 | MFWVLVVVGGVLACYSLLVTVAFIIFWV | CD28 transmembrane domain |
| 298 | KRGRKKLLYIFKQPFMRPVQTTQEEDGCSCRFPEEEEGGCEL | 4-1BB costimulatory domain |
| 299 | RSKRSRLLHSDYMNMTPRRPGPTRKHYQPYAPPRDFAAYRS | CD28 costimulatory domain |
| 300 | RVKFSRSADAPAYQQGQNQLYNELNLGRREEYDVLDKRRGRDPE MGGKPRRKNPQEGLYNELQKDKMAEAYSEIGMKGERRRGKGHD GLYQGLSTATKDTYDALHMQALPPR | CD3ζ signaling domain |
| 301 | RVKFSRSADAPAYKQGQNQLYNELNLGRREEYDVLDKRRGRDPE MGGKPRRKNPQEGLYNELQKDKMAEAYSEIGMKGERRRGKGHD GLYQGLSTATKDTYDALHMQALPPR | CD3ζ signaling domain (with Q to K mutation at position 14) |
| 302 | DIQMTQTTSSLSASLGDRVTISCRASQDISKYLNWYQQKPDGTVKL LIYHTSRLHSGVPSRFSGSGSGTDYSLTISNLEQEDIATYFCQQGNT LPYTFGGGTKLEITGSTSGSGKPGSGEGSTKGEVKLQESGPGLVAP SQSLSVTCTVSGVSLPDYGVSWIRQPPRKGLEWLGVIWGSETTYY NSALKSRLTIIKDNSKSQVFLKMNSLQTDDTAIYYCAKHYYYGGS YAMDYWGQGTSVTVSS | Anti-CD19 FMC63 scFv entire sequence, with Whitlow linker |
| 303 | DIQMTQTTSSLSASLGDRVTISCRASQDISKYLNWYQQKPDGTVKL LIYHTSRLHSGVPSRFSGSGSGTDYSLTISNLEQEDIATYFCQQGNT LPYTFGGGTKLEIT | Anti-CD19 FMC63 scFv light chain variable region |
| 304 | QDISKY | Anti-CD19 FMC63 scFv light chain CDR1 |
| 305 | HTS | Anti-CD19 FMC63 scFv light chain CDR2 |
| 306 | QQGNTLPYT | Anti-CD19 FMC63 scFv light chain CDR3 |
| 307 | GSTSGSGKPGSGEGSTKG | Whitlow linker |
| 308 | EVKLQESGPGLVAPSQSLSVTCTVSGVSLPDYGVSWIRQPPRKGLE WLGVIWGSETTYYNSALKSRLTIIKDNSKSQVFLKMNSLQTDDTAI YYCAKHYYYGGSYAMDYWGQGTSVTVSS | Anti-CD19 FMC63 scFv heavy chain variable region |
| 309 | GVSLPDYG | Anti-CD19 FMC63 scFv |

| | | |
|---|---|---|
| | | heavy chain CDR1 |
| 310 | IWGSETT | Anti-CD19 FMC63 scFv heavy chain CDR2 |
| 311 | AKHYYYGGSYAMDY | Anti-CD19 FMC63 scFv heavy chain CDR3 |
| 312 | DIQMTQTTSSLSASLGDRVTISCRASQDISKYLNWYQQKPDGTVKL LIYHTSRLHSGVPSRFSGSGSGTDYSLTISNLEQEDIATYFCQQGNT LPYTFGGGTKLEITGGGGSGGGGSGGGGSEVKLQESGPGLVAPSQ SLSVTCTVSGVSLPDYGVSWIRQPPRKGLEWLGVIWGSETTYYNS ALKSRLTIIKDNSKSQVFLKMNSLQTDDTAIYYCAKHYYYGGSYA MDYWGQGTSVTVSS | Anti-CD19 FMC63 scFv entire sequence, with 3x$G_4S$ linker |
| 313 | GGGGSGGGGSGGGGS | 3x$G_4S$ linker |
| 314 | atggccttaccagtgaccgccttgctcctgccgctggccttgctgctccacgccgccaggccggacatcc agatgacacagactacatcctccctgtctgcctctctgggagacagagtcaccatcagttgcagggcaag tcaggacattagtaaatatttaaattggtatcagcagaaaccagatggaactgttaaactcctgatctaccat acatcaagattacactcaggagtcccatcaaggttcagtggcagtgggtctggaacagattattctctcacc attagcaacctggagcaagaagatattgccacttacttttgccaacagggtaatacgcttccgtacacgttc ggaggggggaccaagctggagatcacaggctccacctctggatccggcaagcccggatctggcgagg gatccaccaagggcgaggtgaaactgcaggagtcaggacctggcctggtggcgccctcacagagcct gtccgtcacatgcactgtctcaggggtctcattacccgactatggtgtaagctggattcgccagcctccac gaaagggtctggagtggctgggagtaatatggggtagtgaaaccacatactataattcagctctcaaatcc agactgaccatcatcaaggacaactccaagagccaagttttcttaaaaatgaacagtctgcaaactgatga cacagccatttactactgtgccaaacattattactacggtggtagctatgctatggactactggggccaagg aacctcagtcaccgtctcctcaaccacgacgccagcgccgcgaccaccaacaccggcgcccaccatcg cgtcgcagcccctgtccctgcgcccagaggcgtgccggccagcggcgggggggcgcagtgcacacga gggggctggacttcgcctgtgatatctacatctgggcgcccttggccgggacttgtggggtccttctcctgt cactggttatcaccctttactgcaaacggggcagaaagaaactcctgtatatattcaaacaaccatttatgag accagtacaaactactcaagaggaagatggctgtagctgccgatttccagaagaagaagaaggaggatg tgaactgagagtgaagttcagcaggagcgcagacgcccccgcgtaccagcagggccagaaccagctc tataacgagctcaatctaggacgaagagaggagtacgatgttttggacaagagacgtggccgggaccct gagatgggggggaaagccgagaaggaagaaccctcaggaaggcctgtacaatgaactgcagaaagata agatggcggaggcctacagtgagattgggatgaaaggcgagcgccggaggggcaaggggcacgatg gcctttaccagggtctcagtacagccaccaaggacacctacgacgcccttcacatgcaggccctgcccc ctcgc | Exemplary CD19 CAR nucleotide sequence |
| 315 | MALPVTALLLPLALLLHAARPDIQMTQTTSSLSASLGDRVTISCRA SQDISKYLNWYQQKPDGTVKLLIYHTSRLHSGVPSRFSGSGSGTDY SLTISNLEQEDIATYFCQQGNTLPYTFGGGTKLEITGSTSGSGKPGS GEGSTKGEVKLQESGPGLVAPSQSLSVTCTVSGVSLPDYGVSWIR QPPRKGLEWLGVIWGSETTYYNSALKSRLTIIKDNSKSQVFLKMNS LQTDDTAIYYCAKHYYYGGSYAMDYWGQGTSVTVSSTTTPAPRP PTPAPTIASQPLSLRPEACRPAAGGAVHTRGLDFACDIYIWAPLAG TCGVLLLSLVITLYCKRGRKKLLYIFKQPFMRPVQTTQEEDGCSCR FPEEEEGGCELRVKFSRSADAPAYQQGQNQLYNELNLGRREEYDV LDKRRGRDPEMGGKPRRKNPQEGLYNELQKDKMAEAYSEIGMK GERRRGKGHDGLYQGLSTATKDTYDALHMQALPPR | Exemplary CD19 CAR amino acid sequence |
| 316 | atggccttaccagtgaccgccttgctcctgccgctggccttgctgctccacgccgccaggccggacatcc agatgacacagactacatcctccctgtctgcctctctgggagacagagtcaccatcagttgcagggcaag tcaggacattagtaaatatttaaattggtatcagcagaaaccagatggaactgttaaactcctgatctaccat | Tisagenlecleucel CD19 CAR |

| | | |
|---|---|---|
| | acatcaagattacactcaggagtcccatcaaggttcagtggcagtgggtctggaacagattattctctcacc attagcaacctggagcaagaagatattgccacttacttttgccaacagggtaatacgcttccgtacacgttc ggaggggggaccaagctggagatcacaggtggcggtggctcgggcggtggtgggtcgggtggcggc ggatctgaggtgaaactgcaggagtcaggacctggcctggtggcgccctcacagagcctgtccgtcaca tgcactgtctcaggggtctcattacccgactatggtgtaagctggattcgccagcctccacgaaagggtct ggagtggctgggagtaatatggggtagtgaaaccacatactataattcagctctcaaatccagactgacca tcatcaaggacaactccaagagccaagttttcttaaaaatgaacagtctgcaaactgatgacacagccattt actactgtgccaaacattattactacggtggtagctatgctatggactactggggccaaggaacctcagtca ccgtctcctcaaccacgacgccagcgccgcgaccaccaacaccggcgcccaccatcgcgtcgcagcc cctgtccctgcgcccagaggcgtgccggccagcggcgggggcgcagtgcacacgagggggctgg acttcgcctgtgatatctacatctgggcgcccttggccgggacttgtggggtccttctcctgtcactggttat caccctttactgcaaacggggcagaaagaaactcctgtatatattcaaacaaccatttatgagaccagtaca aactactcaagaggaagatggctgtagctgccgatttccagaagaagaagaaggaggatgtgaactgag agtgaagttcagcaggagcgcagacgccccgcgtacaagcagggccagaaccagctctataacgag ctcaatctaggacgaagagaggagtacgatgttttggacaagagacgtggccgggaccctgagatggg gggaaagccgagaaggaagaaccctcaggaaggcctgtacaatgaactgcagaaagataagatggcg gaggcctacagtgagattgggatgaaaggcgagcgccggaggggcaaggggcacgatggcctttacc agggtctcagtacagccaccaaggacacctacgacgcccttcacatgcaggccctgcccctcgc | nucleotide sequence |
| 317 | MALPVTALLLPLALLLHAARPDIQMTQTTSSLSASLGDRVTISCRA SQDISKYLNWYQQKPDGTVKLLIYHTSRLHSGVPSRFSGSGSGTDY SLTISNLEQEDIATYFCQQGNTLPYTFGGGTKLEITGGGGSGGGGS GGGGSEVKLQESGPGLVAPSQSLSVTCTVSGVSLPDYGVSWIRQPP RKGLEWLGVIWGSETTYYNSALKSRLTIIKDNSKSQVFLKMNSLQ TDDTAIYYCAKHYYYGGSYAMDYWGQGTSVTVSSTTTPAPRPPT PAPTIASQPLSLRPEACRPAAGGAVHTRGLDFACDIYIWAPLAGTC GVLLLSLVITLYCKRGRKKLLYIFKQPFMRPVQTTQEEDGCSCRFP EEEEGGCELRVKFSRSADAPAYKQGQNQLYNELNLGRREEYDVL DKRRGRDPEMGGKPRRKNPQEGLYNELQKDKMAEAYSEIGMKG ERRRGKGHDGLYQGLSTATKDTYDALHMQALPPR | Tisagenlecleucel CD19 CAR amino acid sequence |
| 318 | atgctgctgctggtgaccagcctgctgctgtgcgagctgccccaccccgcctttctgctgatccccgacat ccagatgacccagaccacctccagcctgagcgccagcctgggcgaccgggtgaccatcagctgccgg gccagccaggacatcagcaagtacctgaactggtatcagcagaagcccgacggcaccgtcaagctgct gatctaccacaccagccggctgcacagcggcgtgcccagccggtttagcggcagcggctccggcacc gactacagcctgaccatctccaacctggaacaggaagatatcgccacctacttttgccagcagggcaaca cactgccctacacctttggcggcggaacaaagctggaaatcaccggcagcacctccggcagcggcaag cctggcagcggcgagggcagcaccaagggcgaggtgaagctgcaggaaagcggccctggcctggtg gcccccagccagagcctgagcgtgacctgcaccgtgagcggcgtgagcctgcccgactacggcgtga gctggatccggcagccccccaggaagggcctggaatggctgggcgtgatctggggcagcgagaccac ctactacaacagcgccctgaagagccggctgaccatcatcaaggacaacagcaagagccaggtgttcct gaagatgaacagcctgcagaccgacgacaccgccatctactactgcgccaagcactactactacggcg gcagctacgccatggactactggggccagggcaccagcgtgaccgtgagcagcgaatctaagtacgg accgccctgccccccttgccctatgttctgggtgctggtggtggtcggaggcgtgctggcctgctacagc ctgctggtcaccgtggccttcatcatcttttgggtgaaacggggcagaaagaaactcctgtatatattcaaa caaccatttatgagaccagtacaaactactcaagaggaagatggctgtagctgccgatttccagaagaag aagaaggaggatgtgaactgcgggtgaagttcagcagaagcgccgacgcccctgcctaccagcaggg ccagaatcagctgtacaacgagctgaacctgggcagaagggaagagtacgacgtcctggataagcgga gaggccgggaccctgagatgggcggcaagcctcggcggaagaacccccaggaaggcctgtataacg aactgcagaaagacaagatggccgaggcctacagcgagatcggcatgaagggcgagcggaggcgg ggcaagggccacgacggcctgtatcagggcctgtccaccgccaccaaggatacctacgacgccctgca catgcaggccctgcccccaagg | Lisocabtagene maraleucel CD19 CAR nucleotide sequence |
| 319 | MLLLVTSLLLCELPHPAFLLIPDIQMTQTTSSLSASLGDRVTISCRAS QDISKYLNWYQQKPDGTVKLLIYHTSRLHSGVPSRFSGSGSGTDYS LTISNLEQEDIATYFCQQGNTLPYTFGGGTKLEITGSTSGSGKPGSG | Lisocabtagene maraleucel CD19 |

| | | |
|---|---|---|
| | EGSTKGEVKLQESGPGLVAPSQSLSVTCTVSGVSLPDYGVSWIRQP PRKGLEWLGVIWGSETTYYNSALKSRLTIIKDNSKSQVFLKMNSL QTDDTAIYYCAKHYYYGGSYAMDYWGQGTSVTVSSESKYGPPCP PCPMFWVLVVVGGVLACYSLLVTVAFIIFWVKRGRKKLLYIFKQP FMRPVQTTQEEDGCSCRFPEEEEGGCELRVKFSRSADAPAYQQGQ NQLYNELNLGRREEYDVLDKRRGRDPEMGGKPRRKNPQEGLYNE LQKDKMAEAYSEIGMKGERRRGKGHDGLYQGLSTATKDTYDAL HMQALPPR | CAR amino acid sequence |
| 320 | atgcttctcctggtgacaagccttctgctctgtgagttaccacacccagcattcctcctgatcccagacatcc agatgacacagactacatcctccctgtctgcctctctgggagacagagtcaccatcagttgcagggcaag tcaggacattagtaaatatttaaattggtatcagcagaaaccagatggaactgttaaactcctgatctaccat acatcaagattacactcaggagtcccatcaaggttcagtggcagtgggtctggaacagattattctctcacc attagcaacctggagcaagaagatattgccacttacttttgccaacagggtaatacgcttccgtacacgttc ggaggggggactaagttggaaataacaggctccacctctggatccggcaagcccggatctggcgagg gatccaccaagggcgaggtgaaactgcaggagtcaggacctggcctggtggcgccctcacagagcct gtccgtcacatgcactgtctcaggggtctcattacccgactatggtgtaagctggattcgccagcctccac gaaagggtctggagtggctgggagtaatatggggtagtgaaaccacatactataattcagctctcaaatcc agactgaccatcatcaaggacaactccaagagccaagttttcttaaaaatgaacagtctgcaaactgatga cacagccatttactactgtgccaaacattattactacggtggtagctatgctatggactactggggtcaagg aacctcagtcaccgtctcctcagcggccgcaattgaagttatgtatcctcctccttacctagacaatgagaa gagcaatggaaccattatccatgtgaaagggaaacacctttgtccaagtcccctatttcccggaccttctaa gccctttgggtgctggtggtggttgggggagtcctggcttgctatagcttgctagtaacagtggcctttatt attttctgggtgaggagtaagaggagcaggctcctgcacagtgactacatgaacatgactccccgccgcc ccgggcccacccgcaagcattaccagccctatgccccaccacgcgacttcgcagcctatcgctccagag tgaagttcagcaggagcgcagacgcccccgcgtaccagcagggccagaaccagctctataacgagctc aatctaggacgaagagaggagtacgatgttttggacaagagacgtggccgggaccctgagatgggggg aaagccgagaaggaagaaccctcaggaaggcctgtacaatgaactgcagaaagataagatggcggag gcctacagtgagattgggatgaaaggcgagcgccggaggggcaaggggcacgatggccttaccagg gtctcagtacagccaccaaggacacctacgacgcccttcacatgcaggccctgcccctcgc | Axicabtagene ciloleucel CD19 CAR nucleotide sequence |
| 321 | MLLLVTSLLLCELPHPAFLLIPDIQMTQTTSSLSASLGDRVTISCRAS QDISKYLNWYQQKPDGTVKLLIYHTSRLHSGVPSRFSGSGSGTDYS LTISNLEQEDIATYFCQQGNTLPYTFGGGTKLEITGSTSGSGKPGSG EGSTKGEVKLQESGPGLVAPSQSLSVTCTVSGVSLPDYGVSWIRQP PRKGLEWLGVIWGSETTYYNSALKSRLTIIKDNSKSQVFLKMNSL QTDDTAIYYCAKHYYYGGSYAMDYWGQGTSVTVSSAAAIEVMY PPPYLDNEKSNGTIIHVKGKHLCPSPLFPGPSKPFWVLVVVGGVLA CYSLLVTVAFIIFWVRSKRSRLLHSDYMNMTPRRPGPTRKHYQPY APPRDFAAYRSRVKFSRSADAPAYQQGQNQLYNELNLGRREEYD VLDKRRGRDPEMGGKPRRKNPQEGLYNELQKDKMAEAYSEIGM KGERRRGKGHDGLYQGLSTATKDTYDALHMQALPPR | Axicabtagene ciloleucel CD19 CAR amino acid sequence |
| 322 | DIVLTQSPAILSASPGEKVTMTCRASSSVNYMDWYQKKPGSSPKP WIYATSNLASGVPARFSGSGSGTSYSLTISRVEAEDAATYYCQQW SFNPPTFGGGTKLEIKGSTSGSGKPGSGEGSTKGEVQLQQSGAELV KPGASVKMSCKASGYTFTSYNMHWVKQTPGQGLEWIGAIYPGNG DTSYNQKFKGKATLTADKSSSTAYMQLSSLTSEDSADYYCARSNY YGSSYWFFDVWGAGTTVTVSS | Anti-CD20 Leu16 scFv entire sequence, with Whitlow linker |
| 323 | DIVLTQSPAILSASPGEKVTMTCRASSSVNYMDWYQKKPGSSPKP WIYATSNLASGVPARFSGSGSGTSYSLTISRVEAEDAATYYCQQW SFNPPTFGGGTKLEIK | Anti-CD20 Leu16 scFv light chain variable region |
| 324 | RASSSVNYMD | Anti-CD20 Leu16 scFv light chain CDR1 |

| | | |
|---|---|---|
| 325 | ATSNLAS | Anti-CD20 Leu16 scFv light chain CDR2 |
| 326 | QQWSFNPPT | Anti-CD20 Leu16 scFv light chain CDR3 |
| 327 | EVQLQQSGAELVKPGASVKMSCKASGYTFTSYNMHWVKQTPGQ GLEWIGAIYPGNGDTSYNQKFKGKATLTADKSSSTAYMQLSSLTS EDSADYYCARSNYYGSSYWFFDVWGAGTTVTVSS | Anti-CD20 Leu16 scFv heavy chain |
| 328 | SYNMH | Anti-CD20 Leu16 scFv heavy chain CDR1 |
| 329 | AIYPGNGDTSYNQKFKG | Anti-CD20 Leu16 scFv heavy chain CDR2 |
| 330 | SNYYGSSYWFFDV | Anti-CD20 Leu16 scFv heavy chain CDR3 |
| 331 | QVQLQQSGPGLVKPSQTLSLTCAISGDSVSSNSAAWNWIRQSPSRG LEWLGRTYYRSKWYNDYAVSVKSRITINPDTSKNQFSLQLNSVTP EDTAVYYCAREVTGDLEDAFDIWGQGTMVTVSSGGGGSGGGGSG GGGSDIQMTQSPSSLSASVGDRVTITCRASQTIWSYLNWYQQRPG KAPNLLIYAASSLQSGVPSRFSGRGSGTDFTLTISSLQAEDFATYYC QQSYSIPQTFGQGTKLEIK | Anti-CD22 m971 scFv entire sequence, with 3x$G_4S$ linker |
| 332 | QVQLQQSGPGLVKPSQTLSLTCAISGDSVSSNSAAWNWIRQSPSRG LEWLGRTYYRSKWYNDYAVSVKSRITINPDTSKNQFSLQLNSVTP EDTAVYYCAREVTGDLEDAFDIWGQGTMVTVSS | Anti-CD22 m971 scFv heavy chain variable region |
| 333 | GDSVSSNSAA | Anti-CD22 m971 scFv heavy chain CDR1 |
| 334 | TYYRSKWYN | Anti-CD22 m971 scFv heavy chain CDR2 |
| 335 | AREVTGDLEDAFDI | Anti-CD22 m971 scFv heavy chain CDR3 |
| 336 | DIQMTQSPSSLSASVGDRVTITCRASQTIWSYLNWYQQRPGKAPN LLIYAASSLQSGVPSRFSGRGSGTDFTLTISSLQAEDFATYYCQQSY SIPQTFGQGTKLEIK | Anti-CD22 m971 scFv light chain |
| 337 | QTIWSY | Anti-CD22 m971 scFv light chain CDR1 |
| 338 | AAS | Anti-CD22 m971 scFv light chain CDR2 |
| 339 | QQSYSIPQT | Anti-CD22 m971 scFv light chain CDR3 |
| 340 | QVQLQQSGPGMVKPSQTLSLTCAISGDSVSSNSVAWNWIRQSPSR GLEWLGRTYYRSTWYNDYAVSMKSRITINPDTNKNQFSLQLNSVT | Anti-CD22 m971-L7 scFv |

| | | |
|---|---|---|
| | PEDTAVYYCAREVTGDLEDAFDIWGQGTMVTVSSGGGGSGGGGS GGGGSDIQMIQSPSSLSASVGDRVTITCRASQTIWSYLNWYRQRPG EAPNLLIYAASSLQSGVPSRFSGRGSGTDFTLTISSLQAEDFATYYC QQSYSIPQTFGQGTKLEIK | entire sequence, with 3x$G_4S$ linker |
| 341 | QVQLQQSGPGMVKPSQTLSLTCAISGDSVSSNSVAWNWIRQSPSR GLEWLGRTYYRSTWYNDYAVSMKSRITINPDTNKNQFSLQLNSVT PEDTAVYYCAREVTGDLEDAFDIWGQGTMVTVSS | Anti-CD22 m971-L7 scFv heavy chain variable region |
| 342 | GDSVSSNSVA | Anti-CD22 m971-L7 scFv heavy chain CDR1 |
| 343 | TYYRSTWYN | Anti-CD22 m971-L7 scFv heavy chain CDR2 |
| 344 | AREVTGDLEDAFDI | Anti-CD22 m971-L7 scFv heavy chain CDR3 |
| 345 | DIQMIQSPSSLSASVGDRVTITCRASQTIWSYLNWYRQRPGEAPNL LIYAASSLQSGVPSRFSGRGSGTDFTLTISSLQAEDFATYYCQQSYS IPQTFGQGTKLEIK | Anti-CD22 m971-L7 scFv light chain variable region |
| 346 | QTIWSY | Anti-CD22 m971-L7 scFv light chain CDR1 |
| 347 | AAS | Anti-CD22 m971-L7 scFv light chain CDR2 |
| 348 | QQSYSIPQT | Anti-CD22 m971-L7 scFv light chain CDR3 |
| 349 | DIVLTQSPASLAMSLGKRATISCRASESVSVIGAHLIHWYQQKPGQ PPKLLIYLASNLETGVPARFSGSGSGTDFTLTIDPVEEDDVAIYSCL QSRIFPRTFGGGTKLEIKGSTSGSGKPGSGEGSTKGQIQLVQSGPEL KKPGETVKISCKASGYTFTDYSINWVKRAPGKGLKWMGWINTET REPAYAYDFRGRFAFSLETSASTAYLQINNLKYEDTATYFCALDYS YAMDYWGQGTSVTVSS | Anti-BCMA C11D5.3 scFv entire sequence, with Whitlow linker |
| 350 | DIVLTQSPASLAMSLGKRATISCRASESVSVIGAHLIHWYQQKPGQ PPKLLIYLASNLETGVPARFSGSGSGTDFTLTIDPVEEDDVAIYSCL QSRIFPRTFGGGTKLEIK | Anti-BCMA C11D5.3 scFv light chain variable region |
| 351 | RASESVSVIGAHLIH | Anti-BCMA C11D5.3 scFv light chain CDR1 |
| 352 | LASNLET | Anti-BCMA C11D5.3 scFv light chain CDR2 |
| 353 | LQSRIFPRT | Anti-BCMA C11D5.3 scFv light chain CDR3 |

| | | |
|---|---|---|
| 354 | QIQLVQSGPELKKPGETVKISCKASGYTFTDYSINWVKRAPGKGLK WMGWINTETREPAYAYDFRGRFAFSLETSASTAYLQINNLKYEDT ATYFCALDYSYAMDYWGQGTSVTVSS | Anti-BCMA C11D5.3 scFv heavy chain variable region |
| 355 | DYSIN | Anti-BCMA C11D5.3 scFv heavy chain CDR1 |
| 356 | WINTETREPAYAYDFRG | Anti-BCMA C11D5.3 scFv heavy chain CDR2 |
| 357 | DYSYAMDY | Anti-BCMA C11D5.3 scFv heavy chain CDR3 |
| 358 | DIVLTQSPPSLAMSLGKRATISCRASESVTILGSHLIYWYQQKPGQP PTLLIQLASNVQTGVPARFSGSGSRTDFTLTIDPVEEDDVAVYYCL QSRTIPRTFGGGTKLEIKGSTSGSGKPGSGEGSTKGQIQLVQSGPEL KKPGETVKISCKASGYTFRHYSMNWVKQAPGKGLKWMGRINTES GVPIYADDFKGRFAFSVETSASTAYLVINNLKDEDTASYFCSNDYL YSLDFWGQGTALTVSS | Anti-BCMA C12A3.2 scFv entire sequence, with Whitlow linker |
| 359 | DIVLTQSPPSLAMSLGKRATISCRASESVTILGSHLIYWYQQKPGQP PTLLIQLASNVQTGVPARFSGSGSRTDFTLTIDPVEEDDVAVYYCL QSRTIPRTFGGGTKLEIK | Anti-BCMA C12A3.2 scFv light chain variable region |
| 360 | RASESVTILGSHLIY | Anti-BCMA C12A3.2 scFv light chain CDR1 |
| 361 | LASNVQT | Anti-BCMA C12A3.2 scFv light chain CDR2 |
| 362 | LQSRTIPRT | Anti-BCMA C12A3.2 scFv light chain CDR3 |
| 363 | QIQLVQSGPELKKPGETVKISCKASGYTFRHYSMNWVKQAPGKGL KWMGRINTESGVPIYADDFKGRFAFSVETSASTAYLVINNLKDED TASYFCSNDYLYSLDFWGQGTALTVSS | Anti-BCMA C12A3.2 scFv heavy chain variable region |
| 364 | HYSMN | Anti-BCMA C12A3.2 scFv heavy chain CDR1 |
| 365 | RINTESGVPIYADDFKG | Anti-BCMA C12A3.2 scFv heavy chain CDR2 |
| 366 | DYLYSLDF | Anti-BCMA C12A3.2 scFv heavy chain CDR3 |

| | | |
|---|---|---|
| 367 | EVQLLESGGGLVQPGGSLRLSCAASGFTFSSYAMSWVRQAPGKGL EWVSSISGSGDYIYYADSVKGRFTISRDISKNTLYLQMNSLRAEDT AVYYCAKEGTGANSSLADYRGQGTLVTVSS | Anti-BCMA FHVH33 entire sequence |
| 368 | GFTFSSYA | Anti-BCMA FHVH33 CDR1 |
| 369 | ISGSGDYI | Anti-BCMA FHVH33 CDR2 |
| 370 | AKEGTGANSSLADY | Anti-BCMA FHVH33 CDR3 |
| 371 | DIQMTQSPSSLSASVGDRVTITCRASQSISSYLNWYQQKPGKAPKL LIYAASSLQSGVPSRFSGSGSGTDFTLTISSLQPEDFATYYCQQKYD LLTFGGGTKVEIKGSTSGSGKPGSGEGSTKGQLQLQESGPGLVKPS ETLSLTCTVSGGSISSSSYYWGWIRQPPGKGLEWIGSISYSGSTYYN PSLKSRVTISVDTSKNQFSLKLSSVTAADTAVYYCARDRGDTILDV WGQGTMVTVSS | Anti-BCMA CT103A scFv entire sequence, with Whitlow linker |
| 372 | DIQMTQSPSSLSASVGDRVTITCRASQSISSYLNWYQQKPGKAPKL LIYAASSLQSGVPSRFSGSGSGTDFTLTISSLQPEDFATYYCQQKYD LLTFGGGTKVEIK | Anti-BCMA CT103A scFv light chain variable region |
| 373 | QSISSY | Anti-BCMA CT103A scFv light chain CDR1 |
| 374 | AAS | Anti-BCMA CT103A scFv light chain CDR2 |
| 375 | QQKYDLLT | Anti-BCMA CT103A scFv light chain CDR3 |
| 376 | QLQLQESGPGLVKPSETLSLTCTVSGGSISSSSYYWGWIRQPPGKG LEWIGSISYSGSTYYNPSLKSRVTISVDTSKNQFSLKLSSVTAADTA VYYCARDRGDTILDVWGQGTMVTVSS | Anti-BCMA CT103A scFv heavy chain variable region |
| 377 | GGSISSSSYY | Anti-BCMA CT103A scFv heavy chain CDR1 |
| 378 | ISYSGST | Anti-BCMA CT103A scFv heavy chain CDR2 |
| 379 | ARDRGDTILDV | Anti-BCMA CT103A scFv heavy chain CDR3 |
| 380 | atggccttaccagtgaccgccttgctcctgccgctggccttgctgctccacgccgccaggccggacatcc agatgacccagtctccatcctccctgtctgcatctgtaggagacagagtcaccatcacttgccgggcaagt cagagcattagcagctatttaaattggtatcagcagaaaccagggaaagcccctaagctcctgatctatgc tgcatccagtttgcaaagtggggtcccatcaaggttcagtggcagtggatctgggacagatttcactctcac catcagcagtctgcaacctgaagattttgcaacttactactgtcagcaaaaatacgacctcctcacttttggc ggagggaccaaggttgagatcaaaggcagcaccagcggctccggcaagcctggctctggcgagggc agcacaaagggacagctgcagctgcaggagtcgggcccaggactggtgaagccttcggagaccctgt ccctcacctgcactgtctctggtggctccatcagcagtagtagttactactggggctggatccgccagccc ccagggaaggggctggagtggattgggagtatctcctatagtgggagcacctactacaacccgtccctc | Exemplary BCMA CAR nucleotide sequence |

| | | |
|---|---|---|
| | aagagtcgagtcaccatatccgtagacacgtccaagaaccagttctccctgaagctgagttctgtgaccgc<br>cgcagacacggcggtgtactactgcgccagagatcgtggagacaccatactagacgtatggggtcagg<br>gtacaatggtcaccgtcagctcattcgtgcccgtgttcctgcccgccaaacctaccaccacccctgcccct<br>agacctcccaccccagccccaacaatcgccagccagcctctgtctctgcggcccgaagcctgtagacct<br>gctgccggcggagccgtgcacaccagaggcctggacttcgcctgcgacatctacatctgggcccctctg<br>gccggcacctgtggcgtgctgctgctgagcctggtgatcaccctgtactgcaaccaccggaacaaacgg<br>ggcagaaagaaactcctgtatatattcaaacaaccatttatgagaccagtacaaactactcaagaggaaga<br>tggctgtagctgccgatttccagaagaagaagaaggaggatgtgaactgagagtgaagttcagcagatc<br>cgccgacgcccctgcctaccagcagggacagaaccagctgtacaacgagctgaacctgggcagacgg<br>gaagagtacgacgtgctggacaagcggagaggccgggaccccgagatgggcggaaagcccagacg<br>gaagaacccccaggaaggcctgtataacgaactgcagaaagacaagatggccgaggcctacagcgag<br>atcggcatgaagggcgagcggaggcgcggcaagggccacgatggcctgtaccagggcctgagcacc<br>gccaccaaggacacctacgacgccctgcacatgcaggccctgcccccaga | |
| 381 | MALPVTALLLPLALLLHAARPDIQMTQSPSSLSASVGDRVTITCRA<br>SQSISSYLNWYQQKPGKAPKLLIYAASSLQSGVPSRFSGSGSGTDF<br>TLTISSLQPEDFATYYCQQKYDLLTFGGGTKVEIKGSTSGSGKPGS<br>GEGSTKGQLQLQESGPGLVKPSETLSLTCTVSGGSISSSSYYWGWI<br>RQPPGKGLEWIGSISYSGSTYYNPSLKSRVTISVDTSKNQFSLKLSS<br>VTAADTAVYYCARDRGDTILDVWGQGTMVTVSSFVPVFLPAKPT<br>TTPAPRPPTPAPTIASQPLSLRPEACRPAAGGAVHTRGLDFACDIYI<br>WAPLAGTCGVLLLSLVITLYCNHRNKRGRKKLLYIFKQPFMRPVQ<br>TTQEEDGCSCRFPEEEEGGCELRVKFSRSADAPAYQQGQNQLYNE<br>LNLGRREEYDVLDKRRGRDPEMGGKPRRKNPQEGLYNELQKDK<br>MAEAYSEIGMKGERRRGKGHDGLYQGLSTATKDTYDALHMQAL<br>PPR | Exemplary BCMA CAR amino acid sequence |
| 382 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQQYTRSTDNQA<br>VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK<br>ISQSTASINENVNEKCKFTLPPLKIHECNISCPNPLPFREYRPQTEGV<br>SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD<br>EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW<br>TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILNSTYWSGSLM<br>MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG<br>DTLYFPAVGFLVRTEFKYNDSNCPITKCQYSKPENCRLSMGIRPNS<br>HYILRSGLLKYNLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF<br>YQASFSWDTMIKFGDVLTVNPLVVNWRNNTVISRPGQSQCPRFNT<br>CPAICAEGVYNDAFLIDRINWISAGVFLDSNATAANPVFTVFKDNE<br>ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK<br>LFAVKIPEQCT | NivG.690 |
| 383 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQNYTRSTDNQA<br>VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK<br>ISQSTASINENVNEKCKFTLPPLKIHECQISCPNPLPFREYRPQTEGV<br>SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD<br>EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW<br>TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILNSTYWSGSLM<br>MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG<br>DTLYFPAVGFLVRTEFKYNDSNCPITKCQYSKPENCRLSMGIRPNS<br>HYILRSGLLKYNLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF<br>YQASFSWDTMIKFGDVLTVNPLVVNWRNNTVISRPGQSQCPRFNT<br>CPAICAEGVYNDAFLIDRINWISAGVFLDSNATAANPVFTVFKDNE<br>ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK<br>LFAVKIPEQCT | NivG.691 |
| 384 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQNYTRSTDNQA<br>VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK | NivG.693 |

| | | |
|---|---|---|
| | ISQSTASINENVNEKCKFTLPPLKIHECNISCPNPLPFREYRPQTEGV<br>SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD<br>EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW<br>TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILQSTYWSGSLM<br>MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG<br>DTLYFPAVGFLVRTEFKYNDSNCPITKCQYSKPENCRLSMGIRPNS<br>HYILRSGLLKYNLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF<br>YQASFSWDTMIKFGDVLTVNPLVVNWRNNTVISRPGQSQCPRFNT<br>CPAICAEGVYNDAFLIDRINWISAGVFLDSNATAANPVFTVFKDNE<br>ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK<br>LFAVKIPEQCT | |
| 385 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQNYTRSTDNQA<br>VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK<br>ISQSTASINENVNEKCKFTLPPLKIHECNISCPNPLPFREYRPQTEGV<br>SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD<br>EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW<br>TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILNSTYWSGSLM<br>MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG<br>DTLYFPAVGFLVRTEFKYQDSNCPITKCQYSKPENCRLSMGIRPNS<br>HYILRSGLLKYNLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF<br>YQASFSWDTMIKFGDVLTVNPLVVNWRNNTVISRPGQSQCPRFNT<br>CPAICAEGVYNDAFLIDRINWISAGVFLDSNATAANPVFTVFKDNE<br>ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK<br>LFAVKIPEQCT | NivG.694 |
| 386 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQNYTRSTDNQA<br>VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK<br>ISQSTASINENVNEKCKFTLPPLKIHECNISCPNPLPFREYRPQTEGV<br>SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD<br>EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW<br>TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILNSTYWSGSLM<br>MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG<br>DTLYFPAVGFLVRTEFKYNDSNCPITKCQYSKPENCRLSMGIRPNS<br>HYILRSGLLKYQLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF<br>YQASFSWDTMIKFGDVLTVNPLVVNWRNNTVISRPGQSQCPRFNT<br>CPAICAEGVYNDAFLIDRINWISAGVFLDSNATAANPVFTVFKDNE<br>ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK<br>LFAVKIPEQCT | NivG.695 |
| 387 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQNYTRSTDNQA<br>VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK<br>ISQSTASINENVNEKCKFTLPPLKIHECNISCPNPLPFREYRPQTEGV<br>SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD<br>EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW<br>TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILNSTYWSGSLM<br>MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG<br>DTLYFPAVGFLVRTEFKYNDSNCPITKCQYSKPENCRLSMGIRPNS<br>HYILRSGLLKYNLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF<br>YQASFSWDTMIKFGDVLTVNPLVVNWRQNTVISRPGQSQCPRFNT<br>CPAICAEGVYNDAFLIDRINWISAGVFLDSNATAANPVFTVFKDNE<br>ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK<br>LFAVKIPEQCT | NivG.696 |
| 388 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQNYTRSTDNQA<br>VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK<br>ISQSTASINENVNEKCKFTLPPLKIHECNISCPNPLPFREYRPQTEGV | NivG.697 |

| | | |
|---|---|---|
| | SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD<br>EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW<br>TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILNSTYWSGSLM<br>MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG<br>DTLYFPAVGFLVRTEFKYNDSNCPITKCQYSKPENCRLSMGIRPNS<br>HYILRSGLLKYNLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF<br>YQASFSWDTMIKFGDVLTVNPLVVNWRNNTVISRPGQSQCPRFNT<br>CPAICAEGVYNDAFLIDRINWISAGVFLDSQATAANPVFTVFKDNE<br>ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK<br>LFAVKIPEQCT | |
| 389 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQNYTRSTDNQA<br>VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK<br>ISQSTASINENVNEKCKFTLPPLKIHECNISCPNPLPFREYRPQTEGV<br>SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD<br>EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW<br>TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILNSTYWSGSLM<br>MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG<br>DTLYFPAVGFLVRTEFKYQDSNCPITKCQYSKPENCRLSMGIRPNS<br>HYILRSGLLKYQLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF<br>YQASFSWDTMIKFGDVLTVNPLVVNWRQNTVISRPGQSQCPRFNT<br>CPAICAEGVYNDAFLIDRINWISAGVFLDSNATAANPVFTVFKDNE<br>ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK<br>LFAVKIPEQCT | NivG.699 |
| 390 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQNYTRSTDNQA<br>VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK<br>ISQSTASINENVNEKCKFTLPPLKIHECNISCPNPLPFREYRPQTEGV<br>SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD<br>EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW<br>TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILNSTYWSGSLM<br>MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG<br>DTLYFPAVGFLVRTEFKYQDSNCPITKCQYSKPENCRLSMGIRPNS<br>HYILRSGLLKYNLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF<br>YQASFSWDTMIKFGDVLTVNPLVVNWRQNTVISRPGQSQCPRFNT<br>CPAICAEGVYNDAFLIDRINWISAGVFLDSNATAANPVFTVFKDNE<br>ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK<br>LFAVKIPEQCT | NivG.700 |
| 391 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQNYTRSTDNQA<br>VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK<br>ISQSTASINENVNEKCKFTLPPLKIHECNISCPNPLPFREYRPQTEGV<br>SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD<br>EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW<br>TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILNSTYWSGSLM<br>MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG<br>DTLYFPAVGFLVRTEFKYQDSNCPITKCQYSKPENCRLSMGIRPNS<br>HYILRSGLLKYNLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF<br>YQASFSWDTMIKFGDVLTVNPLVVNWRNNTVISRPGQSQCPRFNT<br>CPAICAEGVYNDAFLIDRINWISAGVFLDSQATAANPVFTVFKDNE<br>ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK<br>LFAVKIPEQCT | NivG.701 |
| 392 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQQYTRSTDNQA<br>VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK<br>ISQSTASINENVNEKCKFTLPPLKIHECQISCPNPLPFREYRPQTEGV<br>SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD | NivG.702 |

| | | |
|---|---|---|
| | EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW<br>TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILQSTYWSGSLM<br>MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG<br>DTLYFPAVGFLVRTEFKYQDSNCPITKCQYSKPENCRLSMGIRPNS<br>HYILRSGLLKYQLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF<br>YQASFSWDTMIKFGDVLTVNPLVVNWRQNTVISRPGQSQCPRFNT<br>CPAICAEGVYNDAFLIDRINWISAGVFLDSQATAANPVFTVFKDNE<br>ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK<br>LFAVKIPEQCT | |
| 393 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQNYTRSTDNQA<br>VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK<br>ISQSTASINENVNEKCKFTLPPLKIHECNISCPNPLPFREYRPQTEGV<br>SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD<br>EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW<br>TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILNSTYWSGSLM<br>MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG<br>DTLYFPAVGFLVRTEFKYNDSNCPITKCQYSKPENCRLSMGIRPNS<br>HYILRSGLLKYNLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF<br>YQASFSWDTMIKFGDVLTVNPLVVNWRQNTVISRPGQSQCPRFNT<br>CPAICAEGVYNDAFLIDRINWISAGVFLDSQATAANPVFTVFKDNE<br>ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK<br>LFAVKIPEQCT | NivG.740 |
| 394 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQNYTRSTDNQA<br>VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK<br>ISQSTASINENVNEKCKFTLPPLKIHECNISCPNPLPFREYRPQTEGV<br>SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD<br>EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW<br>TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILNSTYWSGSLM<br>MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG<br>DTLYFPAVGFLVRTEFKYNDSNCPITKCQYSKPENCRLSMGIRPNS<br>HYILRSGLLKYQLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF<br>YQASFSWDTMIKFGDVLTVNPLVVNWRNNTVISRPGQSQCPRFNT<br>CPAICAEGVYNDAFLIDRINWISAGVFLDSQATAANPVFTVFKDNE<br>ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK<br>LFAVKIPEQCT | NivG.742 |
| 395 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQNYTRSTDNQA<br>VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK<br>ISQSTASINENVNEKCKFTLPPLKIHECNISCPNPLPFREYRPQTEGV<br>SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD<br>EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW<br>TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILNSTYWSGSLM<br>MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG<br>DTLYFPAVGFLVRTEFKYNDSNCPITKCQYSKPENCRLSMGIRPNS<br>HYILRSGLLKYQLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF<br>YQASFSWDTMIKFGDVLTVNPLVVNWRQNTVISRPGQSQCPRFNT<br>CPAICAEGVYNDAFLIDRINWISAGVFLDSNATAANPVFTVFKDNE<br>ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK<br>LFAVKIPEQCT | NivG.743 |
| 396 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQNYTRSTDNQA<br>VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK<br>ISQSTASINENVNEKCKFTLPPLKIHECNISCPNPLPFREYRPQTEGV<br>SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD<br>EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW | NivG.744 |

| | | |
|---|---|---|
| | TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILNSTYWSGSLM<br>MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG<br>DTLYFPAVGFLVRTEFKYNDSNCPITKCQYSKPENCRLSMGIRPNS<br>HYILRSGLLKYQLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF<br>YQASFSWDTMIKFGDVLTVNPLVVNWRQNTVISRPGQSQCPRFNT<br>CPAICAEGVYNDAFLIDRINWISAGVFLDSQATAANPVFTVFKDNE<br>ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK<br>LFAVKIPEQCT | |
| 397 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQNYTRSTDNQA<br>VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK<br>ISQSTASINENVNEKCKFTLPPLKIHECNISCPNPLPFREYRPQTEGV<br>SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD<br>EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW<br>TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILNSTYWSGSLM<br>MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG<br>DTLYFPAVGFLVRTEFKYQDSNCPITKCQYSKPENCRLSMGIRPNS<br>HYILRSGLLKYNLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF<br>YQASFSWDTMIKFGDVLTVNPLVVNWRQNTVISRPGQSQCPRFNT<br>CPAICAEGVYNDAFLIDRINWISAGVFLDSQATAANPVFTVFKDNE<br>ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK<br>LFAVKIPEQCT | NivG.748 |
| 398 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQNYTRSTDNQA<br>VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK<br>ISQSTASINENVNEKCKFTLPPLKIHECNISCPNPLPFREYRPQTEGV<br>SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD<br>EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW<br>TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILNSTYWSGSLM<br>MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG<br>DTLYFPAVGFLVRTEFKYQDSNCPITKCQYSKPENCRLSMGIRPNS<br>HYILRSGLLKYQLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF<br>YQASFSWDTMIKFGDVLTVNPLVVNWRNNTVISRPGQSQCPRFNT<br>CPAICAEGVYNDAFLIDRINWISAGVFLDSNATAANPVFTVFKDNE<br>ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK<br>LFAVKIPEQCT | NivG.749 |
| 399 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQNYTRSTDNQA<br>VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK<br>ISQSTASINENVNEKCKFTLPPLKIHECNISCPNPLPFREYRPQTEGV<br>SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD<br>EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW<br>TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILNSTYWSGSLM<br>MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG<br>DTLYFPAVGFLVRTEFKYQDSNCPITKCQYSKPENCRLSMGIRPNS<br>HYILRSGLLKYQLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF<br>YQASFSWDTMIKFGDVLTVNPLVVNWRNNTVISRPGQSQCPRFNT<br>CPAICAEGVYNDAFLIDRINWISAGVFLDSQATAANPVFTVFKDNE<br>ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK<br>LFAVKIPEQCT | NivG.750 |
| 400 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQNYTRSTDNQA<br>VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK<br>ISQSTASINENVNEKCKFTLPPLKIHECNISCPNPLPFREYRPQTEGV<br>SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD<br>EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW<br>TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILNSTYWSGSLM | NivG.752 |

| | | |
|---|---|---|
| | MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG<br>DTLYFPAVGFLVRTEFKYQDSNCPITKCQYSKPENCRLSMGIRPNS<br>HYILRSGLLKYQLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF<br>YQASFSWDTMIKFGDVLTVNPLVVNWRQNTVISRPGQSQCPRFNT<br>CPAICAEGVYNDAFLIDRINWISAGVFLDSQATAANPVFTVFKDNE<br>ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK<br>LFAVKIPEQCT | |
| 401 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQNYTRSTDNQA<br>VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK<br>ISQSTASINENVNEKCKFTLPPLKIHECNISCPNPLPFREYRPQTEGV<br>SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD<br>EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW<br>TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILQSTYWSGSLM<br>MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG<br>DTLYFPAVGFLVRTEFKYNDSNCPITKCQYSKPENCRLSMGIRPNS<br>HYILRSGLLKYNLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF<br>YQASFSWDTMIKFGDVLTVNPLVVNWRNNTVISRPGQSQCPRFNT<br>CPAICAEGVYNDAFLIDRINWISAGVFLDSQATAANPVFTVFKDNE<br>ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK<br>LFAVKIPEQCT | NivG.753 |
| 402 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQNYTRSTDNQA<br>VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK<br>ISQSTASINENVNEKCKFTLPPLKIHECNISCPNPLPFREYRPQTEGV<br>SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD<br>EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW<br>TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILQSTYWSGSLM<br>MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG<br>DTLYFPAVGFLVRTEFKYNDSNCPITKCQYSKPENCRLSMGIRPNS<br>HYILRSGLLKYNLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF<br>YQASFSWDTMIKFGDVLTVNPLVVNWRQNTVISRPGQSQCPRFNT<br>CPAICAEGVYNDAFLIDRINWISAGVFLDSNATAANPVFTVFKDNE<br>ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK<br>LFAVKIPEQCT | NivG.754 |
| 403 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQNYTRSTDNQA<br>VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK<br>ISQSTASINENVNEKCKFTLPPLKIHECNISCPNPLPFREYRPQTEGV<br>SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD<br>EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW<br>TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILQSTYWSGSLM<br>MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG<br>DTLYFPAVGFLVRTEFKYNDSNCPITKCQYSKPENCRLSMGIRPNS<br>HYILRSGLLKYNLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF<br>YQASFSWDTMIKFGDVLTVNPLVVNWRQNTVISRPGQSQCPRFNT<br>CPAICAEGVYNDAFLIDRINWISAGVFLDSQATAANPVFTVFKDNE<br>ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK<br>LFAVKIPEQCT | NivG.755 |
| 404 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQNYTRSTDNQA<br>VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK<br>ISQSTASINENVNEKCKFTLPPLKIHECNISCPNPLPFREYRPQTEGV<br>SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD<br>EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW<br>TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILQSTYWSGSLM<br>MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG | NivG.756 |

| | | |
|---|---|---|
| | DTLYFPAVGFLVRTEFKYNDSNCPITKCQYSKPENCRLSMGIRPNS<br>HYILRSGLLKYQLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF<br>YQASFSWDTMIKFGDVLTVNPLVVNWRNNTVISRPGQSQCPRFNT<br>CPAICAEGVYNDAFLIDRINWISAGVFLDSNATAANPVFTVFKDNE<br>ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK<br>LFAVKIPEQCT | |
| 405 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQNYTRSTDNQA<br>VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK<br>ISQSTASINENVNEKCKFTLPPLKIHECNISCPNPLPFREYRPQTEGV<br>SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD<br>EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW<br>TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILQSTYWSGSLM<br>MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG<br>DTLYFPAVGFLVRTEFKYNDSNCPITKCQYSKPENCRLSMGIRPNS<br>HYILRSGLLKYQLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF<br>YQASFSWDTMIKFGDVLTVNPLVVNWRNNTVISRPGQSQCPRFNT<br>CPAICAEGVYNDAFLIDRINWISAGVFLDSQATAANPVFTVFKDNE<br>ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK<br>LFAVKIPEQCT | NivG.757 |
| 406 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQNYTRSTDNQA<br>VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK<br>ISQSTASINENVNEKCKFTLPPLKIHECNISCPNPLPFREYRPQTEGV<br>SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD<br>EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW<br>TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILQSTYWSGSLM<br>MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG<br>DTLYFPAVGFLVRTEFKYNDSNCPITKCQYSKPENCRLSMGIRPNS<br>HYILRSGLLKYQLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF<br>YQASFSWDTMIKFGDVLTVNPLVVNWRQNTVISRPGQSQCPRFNT<br>CPAICAEGVYNDAFLIDRINWISAGVFLDSNATAANPVFTVFKDNE<br>ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK<br>LFAVKIPEQCT | NivG.758 |
| 407 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQNYTRSTDNQA<br>VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK<br>ISQSTASINENVNEKCKFTLPPLKIHECNISCPNPLPFREYRPQTEGV<br>SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD<br>EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW<br>TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILQSTYWSGSLM<br>MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG<br>DTLYFPAVGFLVRTEFKYNDSNCPITKCQYSKPENCRLSMGIRPNS<br>HYILRSGLLKYQLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF<br>YQASFSWDTMIKFGDVLTVNPLVVNWRQNTVISRPGQSQCPRFNT<br>CPAICAEGVYNDAFLIDRINWISAGVFLDSQATAANPVFTVFKDNE<br>ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK<br>LFAVKIPEQCT | NivG.759 |
| 408 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQNYTRSTDNQA<br>VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK<br>ISQSTASINENVNEKCKFTLPPLKIHECNISCPNPLPFREYRPQTEGV<br>SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD<br>EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW<br>TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILQSTYWSGSLM<br>MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG<br>DTLYFPAVGFLVRTEFKYQDSNCPITKCQYSKPENCRLSMGIRPNS | NivG.760 |

| | | |
|---|---|---|
| | HYILRSGLLKYNLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF<br>YQASFSWDTMIKFGDVLTVNPLVVNWRNNTVISRPGQSQCPRFNT<br>CPAICAEGVYNDAFLIDRINWISAGVFLDSNATAANPVFTVFKDNE<br>ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK<br>LFAVKIPEQCT | |
| 409 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQNYTRSTDNQA<br>VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK<br>ISQSTASINENVNEKCKFTLPPLKIHECNISCPNPLPFREYRPQTEGV<br>SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD<br>EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW<br>TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILQSTYWSGSLM<br>MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG<br>DTLYFPAVGFLVRTEFKYQDSNCPITKCQYSKPENCRLSMGIRPNS<br>HYILRSGLLKYNLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF<br>YQASFSWDTMIKFGDVLTVNPLVVNWRNNTVISRPGQSQCPRFNT<br>CPAICAEGVYNDAFLIDRINWISAGVFLDSQATAANPVFTVFKDNE<br>ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK<br>LFAVKIPEQCT | NivG.761 |
| 410 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQNYTRSTDNQA<br>VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK<br>ISQSTASINENVNEKCKFTLPPLKIHECNISCPNPLPFREYRPQTEGV<br>SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD<br>EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW<br>TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILQSTYWSGSLM<br>MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG<br>DTLYFPAVGFLVRTEFKYQDSNCPITKCQYSKPENCRLSMGIRPNS<br>HYILRSGLLKYNLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF<br>YQASFSWDTMIKFGDVLTVNPLVVNWRQNTVISRPGQSQCPRFNT<br>CPAICAEGVYNDAFLIDRINWISAGVFLDSNATAANPVFTVFKDNE<br>ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK<br>LFAVKIPEQCT | NivG.762 |
| 411 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQNYTRSTDNQA<br>VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK<br>ISQSTASINENVNEKCKFTLPPLKIHECNISCPNPLPFREYRPQTEGV<br>SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD<br>EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW<br>TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILQSTYWSGSLM<br>MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG<br>DTLYFPAVGFLVRTEFKYQDSNCPITKCQYSKPENCRLSMGIRPNS<br>HYILRSGLLKYNLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF<br>YQASFSWDTMIKFGDVLTVNPLVVNWRQNTVISRPGQSQCPRFNT<br>CPAICAEGVYNDAFLIDRINWISAGVFLDSQATAANPVFTVFKDNE<br>ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK<br>LFAVKIPEQCT | NivG.763 |
| 412 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQNYTRSTDNQA<br>VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK<br>ISQSTASINENVNEKCKFTLPPLKIHECNISCPNPLPFREYRPQTEGV<br>SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD<br>EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW<br>TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILQSTYWSGSLM<br>MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG<br>DTLYFPAVGFLVRTEFKYQDSNCPITKCQYSKPENCRLSMGIRPNS<br>HYILRSGLLKYQLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF | NivG.764 |

| | | |
|---|---|---|
| | YQASFSWDTMIKFGDVLTVNPLVVNWRNNTVISRPGQSQCPRFNT<br>CPAICAEGVYNDAFLIDRINWISAGVFLDSNATAANPVFTVFKDNE<br>ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK<br>LFAVKIPEQCT | |
| 413 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQNYTRSTDNQA<br>VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK<br>ISQSTASINENVNEKCKFTLPPLKIHECNISCPNPLPFREYRPQTEGV<br>SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD<br>EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW<br>TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILQSTYWSGSLM<br>MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG<br>DTLYFPAVGFLVRTEFKYQDSNCPITKCQYSKPENCRLSMGIRPNS<br>HYILRSGLLKYQLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF<br>YQASFSWDTMIKFGDVLTVNPLVVNWRNNTVISRPGQSQCPRFNT<br>CPAICAEGVYNDAFLIDRINWISAGVFLDSQATAANPVFTVFKDNE<br>ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK<br>LFAVKIPEQCT | NivG.765 |
| 414 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQNYTRSTDNQA<br>VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK<br>ISQSTASINENVNEKCKFTLPPLKIHECNISCPNPLPFREYRPQTEGV<br>SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD<br>EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW<br>TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILQSTYWSGSLM<br>MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG<br>DTLYFPAVGFLVRTEFKYQDSNCPITKCQYSKPENCRLSMGIRPNS<br>HYILRSGLLKYQLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF<br>YQASFSWDTMIKFGDVLTVNPLVVNWRQNTVISRPGQSQCPRFNT<br>CPAICAEGVYNDAFLIDRINWISAGVFLDSNATAANPVFTVFKDNE<br>ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK<br>LFAVKIPEQCT | NivG.766 |
| 415 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQNYTRSTDNQA<br>VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK<br>ISQSTASINENVNEKCKFTLPPLKIHECNISCPNPLPFREYRPQTEGV<br>SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD<br>EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW<br>TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILQSTYWSGSLM<br>MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG<br>DTLYFPAVGFLVRTEFKYQDSNCPITKCQYSKPENCRLSMGIRPNS<br>HYILRSGLLKYQLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF<br>YQASFSWDTMIKFGDVLTVNPLVVNWRQNTVISRPGQSQCPRFNT<br>CPAICAEGVYNDAFLIDRINWISAGVFLDSQATAANPVFTVFKDNE<br>ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK<br>LFAVKIPEQCT | NivG.767 |
| 416 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQNYTRSTDNQA<br>VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK<br>ISQSTASINENVNEKCKFTLPPLKIHECQISCPNPLPFREYRPQTEGV<br>SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD<br>EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW<br>TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILNSTYWSGSLM<br>MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG<br>DTLYFPAVGFLVRTEFKYNDSNCPITKCQYSKPENCRLSMGIRPNS<br>HYILRSGLLKYNLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF<br>YQASFSWDTMIKFGDVLTVNPLVVNWRNNTVISRPGQSQCPRFNT | NivG.768 |

| | | |
|---|---|---|
| | CPAICAEGVYNDAFLIDRINWISAGVFLDSQATAANPVFTVFKDNE ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK LFAVKIPEQCT | |
| 417 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQNYTRSTDNQA VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK ISQSTASINENVNEKCKFTLPPLKIHECQISCPNPLPFREYRPQTEGV SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILNSTYWSGSLM MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG DTLYFPAVGFLVRTEFKYNDSNCPITKCQYSKPENCRLSMGIRPNS HYILRSGLLKYNLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF YQASFSWDTMIKFGDVLTVNPLVVNWRQNTVISRPGQSQCPRFNT CPAICAEGVYNDAFLIDRINWISAGVFLDSNATAANPVFTVFKDNE ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK LFAVKIPEQCT | NivG.769 |
| 418 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQNYTRSTDNQA VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK ISQSTASINENVNEKCKFTLPPLKIHECQISCPNPLPFREYRPQTEGV SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILNSTYWSGSLM MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG DTLYFPAVGFLVRTEFKYNDSNCPITKCQYSKPENCRLSMGIRPNS HYILRSGLLKYNLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF YQASFSWDTMIKFGDVLTVNPLVVNWRQNTVISRPGQSQCPRFNT CPAICAEGVYNDAFLIDRINWISAGVFLDSQATAANPVFTVFKDNE ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK LFAVKIPEQCT | NivG.770 |
| 419 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQNYTRSTDNQA VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK ISQSTASINENVNEKCKFTLPPLKIHECQISCPNPLPFREYRPQTEGV SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILNSTYWSGSLM MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG DTLYFPAVGFLVRTEFKYNDSNCPITKCQYSKPENCRLSMGIRPNS HYILRSGLLKYQLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF YQASFSWDTMIKFGDVLTVNPLVVNWRNNTVISRPGQSQCPRFNT CPAICAEGVYNDAFLIDRINWISAGVFLDSNATAANPVFTVFKDNE ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK LFAVKIPEQCT | NivG.771 |
| 420 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQNYTRSTDNQA VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK ISQSTASINENVNEKCKFTLPPLKIHECQISCPNPLPFREYRPQTEGV SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILNSTYWSGSLM MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG DTLYFPAVGFLVRTEFKYNDSNCPITKCQYSKPENCRLSMGIRPNS HYILRSGLLKYQLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF YQASFSWDTMIKFGDVLTVNPLVVNWRNNTVISRPGQSQCPRFNT CPAICAEGVYNDAFLIDRINWISAGVFLDSQATAANPVFTVFKDNE | NivG.772 |

| | | |
|---|---|---|
| | ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK LFAVKIPEQCT | |
| 421 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQNYTRSTDNQA VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK ISQSTASINENVNEKCKFTLPPLKIHECQISCPNPLPFREYRPQTEGV SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILNSTYWSGSLM MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG DTLYFPAVGFLVRTEFKYNDSNCPITKCQYSKPENCRLSMGIRPNS HYILRSGLLKYQLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF YQASFSWDTMIKFGDVLTVNPLVVNWRQNTVISRPGQSQCPRFNT CPAICAEGVYNDAFLIDRINWISAGVFLDSNATAANPVFTVFKDNE ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK LFAVKIPEQCT | NivG.773 |
| 422 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQNYTRSTDNQA VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK ISQSTASINENVNEKCKFTLPPLKIHECQISCPNPLPFREYRPQTEGV SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILNSTYWSGSLM MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG DTLYFPAVGFLVRTEFKYNDSNCPITKCQYSKPENCRLSMGIRPNS HYILRSGLLKYQLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF YQASFSWDTMIKFGDVLTVNPLVVNWRQNTVISRPGQSQCPRFNT CPAICAEGVYNDAFLIDRINWISAGVFLDSQATAANPVFTVFKDNE ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK LFAVKIPEQCT | NivG.774 |
| 423 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQNYTRSTDNQA VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK ISQSTASINENVNEKCKFTLPPLKIHECQISCPNPLPFREYRPQTEGV SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILNSTYWSGSLM MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG DTLYFPAVGFLVRTEFKYQDSNCPITKCQYSKPENCRLSMGIRPNS HYILRSGLLKYNLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF YQASFSWDTMIKFGDVLTVNPLVVNWRNNTVISRPGQSQCPRFNT CPAICAEGVYNDAFLIDRINWISAGVFLDSNATAANPVFTVFKDNE ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK LFAVKIPEQCT | NivG.775 |
| 424 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQNYTRSTDNQA VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK ISQSTASINENVNEKCKFTLPPLKIHECQISCPNPLPFREYRPQTEGV SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILNSTYWSGSLM MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG DTLYFPAVGFLVRTEFKYQDSNCPITKCQYSKPENCRLSMGIRPNS HYILRSGLLKYNLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF YQASFSWDTMIKFGDVLTVNPLVVNWRNNTVISRPGQSQCPRFNT CPAICAEGVYNDAFLIDRINWISAGVFLDSQATAANPVFTVFKDNE | NivG.776 |

| | | |
|---|---|---|
| | ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK LFAVKIPEQCT | |
| 425 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQNYTRSTDNQA VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK ISQSTASINENVNEKCKFTLPPLKIHECQISCPNPLPFREYRPQTEGV SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILNSTYWSGSLM MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG DTLYFPAVGFLVRTEFKYQDSNCPITKCQYSKPENCRLSMGIRPNS HYILRSGLLKYNLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF YQASFSWDTMIKFGDVLTVNPLVVNWRQNTVISRPGQSQCPRFNT CPAICAEGVYNDAFLIDRINWISAGVFLDSNATAANPVFTVFKDNE ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK LFAVKIPEQCT | NivG.777 |
| 426 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQNYTRSTDNQA VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK ISQSTASINENVNEKCKFTLPPLKIHECQISCPNPLPFREYRPQTEGV SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILNSTYWSGSLM MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG DTLYFPAVGFLVRTEFKYQDSNCPITKCQYSKPENCRLSMGIRPNS HYILRSGLLKYNLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF YQASFSWDTMIKFGDVLTVNPLVVNWRQNTVISRPGQSQCPRFNT CPAICAEGVYNDAFLIDRINWISAGVFLDSQATAANPVFTVFKDNE ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK LFAVKIPEQCT | NivG.778 |
| 427 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQNYTRSTDNQA VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK ISQSTASINENVNEKCKFTLPPLKIHECQISCPNPLPFREYRPQTEGV SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILNSTYWSGSLM MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG DTLYFPAVGFLVRTEFKYQDSNCPITKCQYSKPENCRLSMGIRPNS HYILRSGLLKYQLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF YQASFSWDTMIKFGDVLTVNPLVVNWRNNTVISRPGQSQCPRFNT CPAICAEGVYNDAFLIDRINWISAGVFLDSNATAANPVFTVFKDNE ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK LFAVKIPEQCT | NivG.779 |
| 428 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQNYTRSTDNQA VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK ISQSTASINENVNEKCKFTLPPLKIHECQISCPNPLPFREYRPQTEGV SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILNSTYWSGSLM MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG DTLYFPAVGFLVRTEFKYQDSNCPITKCQYSKPENCRLSMGIRPNS HYILRSGLLKYQLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF YQASFSWDTMIKFGDVLTVNPLVVNWRNNTVISRPGQSQCPRFNT CPAICAEGVYNDAFLIDRINWISAGVFLDSQATAANPVFTVFKDNE | NivG.780 |

| | | |
|---|---|---|
| | ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK<br>LFAVKIPEQCT | |
| 429 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQNYTRSTDNQA<br>VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK<br>ISQSTASINENVNEKCKFTLPPLKIHECQISCPNPLPFREYRPQTEGV<br>SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD<br>EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW<br>TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILNSTYWSGSLM<br>MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG<br>DTLYFPAVGFLVRTEFKYQDSNCPITKCQYSKPENCRLSMGIRPNS<br>HYILRSGLLKYQLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF<br>YQASFSWDTMIKFGDVLTVNPLVVNWRQNTVISRPGQSQCPRFNT<br>CPAICAEGVYNDAFLIDRINWISAGVFLDSNATAANPVFTVFKDNE<br>ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK<br>LFAVKIPEQCT | NivG.781 |
| 430 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQNYTRSTDNQA<br>VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK<br>ISQSTASINENVNEKCKFTLPPLKIHECQISCPNPLPFREYRPQTEGV<br>SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD<br>EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW<br>TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILNSTYWSGSLM<br>MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG<br>DTLYFPAVGFLVRTEFKYQDSNCPITKCQYSKPENCRLSMGIRPNS<br>HYILRSGLLKYQLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF<br>YQASFSWDTMIKFGDVLTVNPLVVNWRQNTVISRPGQSQCPRFNT<br>CPAICAEGVYNDAFLIDRINWISAGVFLDSQATAANPVFTVFKDNE<br>ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK<br>LFAVKIPEQCT | NivG.782 |
| 431 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQNYTRSTDNQA<br>VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK<br>ISQSTASINENVNEKCKFTLPPLKIHECQISCPNPLPFREYRPQTEGV<br>SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD<br>EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW<br>TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILQSTYWSGSLM<br>MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG<br>DTLYFPAVGFLVRTEFKYNDSNCPITKCQYSKPENCRLSMGIRPNS<br>HYILRSGLLKYNLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF<br>YQASFSWDTMIKFGDVLTVNPLVVNWRNNTVISRPGQSQCPRFNT<br>CPAICAEGVYNDAFLIDRINWISAGVFLDSNATAANPVFTVFKDNE<br>ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK<br>LFAVKIPEQCT | NivG.783 |
| 432 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQNYTRSTDNQA<br>VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK<br>ISQSTASINENVNEKCKFTLPPLKIHECQISCPNPLPFREYRPQTEGV<br>SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD<br>EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW<br>TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILQSTYWSGSLM<br>MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG<br>DTLYFPAVGFLVRTEFKYNDSNCPITKCQYSKPENCRLSMGIRPNS<br>HYILRSGLLKYNLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF<br>YQASFSWDTMIKFGDVLTVNPLVVNWRNNTVISRPGQSQCPRFNT<br>CPAICAEGVYNDAFLIDRINWISAGVFLDSQATAANPVFTVFKDNE | NivG.784 |

| | | |
|---|---|---|
| | ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK<br>LFAVKIPEQCT | |
| 433 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQNYTRSTDNQA<br>VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK<br>ISQSTASINENVNEKCKFTLPPLKIHECQISCPNPLPFREYRPQTEGV<br>SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD<br>EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW<br>TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILQSTYWSGSLM<br>MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG<br>DTLYFPAVGFLVRTEFKYNDSNCPITKCQYSKPENCRLSMGIRPNS<br>HYILRSGLLKYNLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF<br>YQASFSWDTMIKFGDVLTVNPLVVNWRQNTVISRPGQSQCPRFNT<br>CPAICAEGVYNDAFLIDRINWISAGVFLDSNATAANPVFTVFKDNE<br>ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK<br>LFAVKIPEQCT | NivG.785 |
| 434 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQNYTRSTDNQA<br>VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK<br>ISQSTASINENVNEKCKFTLPPLKIHECQISCPNPLPFREYRPQTEGV<br>SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD<br>EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW<br>TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILQSTYWSGSLM<br>MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG<br>DTLYFPAVGFLVRTEFKYNDSNCPITKCQYSKPENCRLSMGIRPNS<br>HYILRSGLLKYNLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF<br>YQASFSWDTMIKFGDVLTVNPLVVNWRQNTVISRPGQSQCPRFNT<br>CPAICAEGVYNDAFLIDRINWISAGVFLDSQATAANPVFTVFKDNE<br>ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK<br>LFAVKIPEQCT | NivG.786 |
| 435 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQNYTRSTDNQA<br>VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK<br>ISQSTASINENVNEKCKFTLPPLKIHECQISCPNPLPFREYRPQTEGV<br>SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD<br>EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW<br>TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILQSTYWSGSLM<br>MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG<br>DTLYFPAVGFLVRTEFKYNDSNCPITKCQYSKPENCRLSMGIRPNS<br>HYILRSGLLKYQLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF<br>YQASFSWDTMIKFGDVLTVNPLVVNWRNNTVISRPGQSQCPRFNT<br>CPAICAEGVYNDAFLIDRINWISAGVFLDSNATAANPVFTVFKDNE<br>ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK<br>LFAVKIPEQCT | NivG.787 |
| 436 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQNYTRSTDNQA<br>VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK<br>ISQSTASINENVNEKCKFTLPPLKIHECQISCPNPLPFREYRPQTEGV<br>SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD<br>EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW<br>TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILQSTYWSGSLM<br>MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG<br>DTLYFPAVGFLVRTEFKYNDSNCPITKCQYSKPENCRLSMGIRPNS<br>HYILRSGLLKYQLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF<br>YQASFSWDTMIKFGDVLTVNPLVVNWRNNTVISRPGQSQCPRFNT<br>CPAICAEGVYNDAFLIDRINWISAGVFLDSQATAANPVFTVFKDNE | NivG.788 |

| | | |
|---|---|---|
| | ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK LFAVKIPEQCT | |
| 437 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQNYTRSTDNQA VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK ISQSTASINENVNEKCKFTLPPLKIHECQISCPNPLPFREYRPQTEGV SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILQSTYWSGSLM MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG DTLYFPAVGFLVRTEFKYNDSNCPITKCQYSKPENCRLSMGIRPNS HYILRSGLLKYQLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF YQASFSWDTMIKFGDVLTVNPLVVNWRQNTVISRPGQSQCPRFNT CPAICAEGVYNDAFLIDRINWISAGVFLDSNATAANPVFTVFKDNE ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK LFAVKIPEQCT | NivG.789 |
| 438 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQNYTRSTDNQA VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK ISQSTASINENVNEKCKFTLPPLKIHECQISCPNPLPFREYRPQTEGV SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILQSTYWSGSLM MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG DTLYFPAVGFLVRTEFKYNDSNCPITKCQYSKPENCRLSMGIRPNS HYILRSGLLKYQLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF YQASFSWDTMIKFGDVLTVNPLVVNWRQNTVISRPGQSQCPRFNT CPAICAEGVYNDAFLIDRINWISAGVFLDSQATAANPVFTVFKDNE ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK LFAVKIPEQCT | NivG.790 |
| 439 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQNYTRSTDNQA VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK ISQSTASINENVNEKCKFTLPPLKIHECQISCPNPLPFREYRPQTEGV SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILQSTYWSGSLM MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG DTLYFPAVGFLVRTEFKYQDSNCPITKCQYSKPENCRLSMGIRPNS HYILRSGLLKYNLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF YQASFSWDTMIKFGDVLTVNPLVVNWRNNTVISRPGQSQCPRFNT CPAICAEGVYNDAFLIDRINWISAGVFLDSNATAANPVFTVFKDNE ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK LFAVKIPEQCT | NivG.791 |
| 440 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQNYTRSTDNQA VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK ISQSTASINENVNEKCKFTLPPLKIHECQISCPNPLPFREYRPQTEGV SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILQSTYWSGSLM MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG DTLYFPAVGFLVRTEFKYQDSNCPITKCQYSKPENCRLSMGIRPNS HYILRSGLLKYNLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF YQASFSWDTMIKFGDVLTVNPLVVNWRNNTVISRPGQSQCPRFNT CPAICAEGVYNDAFLIDRINWISAGVFLDSQATAANPVFTVFKDNE | NivG.792 |

| | | |
|---|---|---|
| | ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK<br>LFAVKIPEQCT | |
| 441 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQNYTRSTDNQA<br>VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK<br>ISQSTASINENVNEKCKFTLPPLKIHECQISCPNPLPFREYRPQTEGV<br>SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD<br>EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW<br>TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILQSTYWSGSLM<br>MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG<br>DTLYFPAVGFLVRTEFKYQDSNCPITKCQYSKPENCRLSMGIRPNS<br>HYILRSGLLKYNLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF<br>YQASFSWDTMIKFGDVLTVNPLVVNWRQNTVISRPGQSQCPRFNT<br>CPAICAEGVYNDAFLIDRINWISAGVFLDSNATAANPVFTVFKDNE<br>ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK<br>LFAVKIPEQCT | NivG.793 |
| 442 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQNYTRSTDNQA<br>VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK<br>ISQSTASINENVNEKCKFTLPPLKIHECQISCPNPLPFREYRPQTEGV<br>SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD<br>EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW<br>TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILQSTYWSGSLM<br>MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG<br>DTLYFPAVGFLVRTEFKYQDSNCPITKCQYSKPENCRLSMGIRPNS<br>HYILRSGLLKYNLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF<br>YQASFSWDTMIKFGDVLTVNPLVVNWRQNTVISRPGQSQCPRFNT<br>CPAICAEGVYNDAFLIDRINWISAGVFLDSQATAANPVFTVFKDNE<br>ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK<br>LFAVKIPEQCT | NivG.794 |
| 443 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQNYTRSTDNQA<br>VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK<br>ISQSTASINENVNEKCKFTLPPLKIHECQISCPNPLPFREYRPQTEGV<br>SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD<br>EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW<br>TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILQSTYWSGSLM<br>MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG<br>DTLYFPAVGFLVRTEFKYQDSNCPITKCQYSKPENCRLSMGIRPNS<br>HYILRSGLLKYQLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF<br>YQASFSWDTMIKFGDVLTVNPLVVNWRNNTVISRPGQSQCPRFNT<br>CPAICAEGVYNDAFLIDRINWISAGVFLDSNATAANPVFTVFKDNE<br>ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK<br>LFAVKIPEQCT | NivG.795 |
| 444 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQNYTRSTDNQA<br>VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK<br>ISQSTASINENVNEKCKFTLPPLKIHECQISCPNPLPFREYRPQTEGV<br>SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD<br>EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW<br>TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILQSTYWSGSLM<br>MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG<br>DTLYFPAVGFLVRTEFKYQDSNCPITKCQYSKPENCRLSMGIRPNS<br>HYILRSGLLKYQLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF<br>YQASFSWDTMIKFGDVLTVNPLVVNWRNNTVISRPGQSQCPRFNT<br>CPAICAEGVYNDAFLIDRINWISAGVFLDSQATAANPVFTVFKDNE | NivG.796 |

| | | |
|---|---|---|
| | ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK LFAVKIPEQCT | |
| 445 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQNYTRSTDNQA VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK ISQSTASINENVNEKCKFTLPPLKIHECQISCPNPLPFREYRPQTEGV SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILQSTYWSGSLM MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG DTLYFPAVGFLVRTEFKYQDSNCPITKCQYSKPENCRLSMGIRPNS HYILRSGLLKYQLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF YQASFSWDTMIKFGDVLTVNPLVVNWRQNTVISRPGQSQCPRFNT CPAICAEGVYNDAFLIDRINWISAGVFLDSNATAANPVFTVFKDNE ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK LFAVKIPEQCT | NivG.797 |
| 446 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQNYTRSTDNQA VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK ISQSTASINENVNEKCKFTLPPLKIHECQISCPNPLPFREYRPQTEGV SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILQSTYWSGSLM MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG DTLYFPAVGFLVRTEFKYQDSNCPITKCQYSKPENCRLSMGIRPNS HYILRSGLLKYQLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF YQASFSWDTMIKFGDVLTVNPLVVNWRQNTVISRPGQSQCPRFNT CPAICAEGVYNDAFLIDRINWISAGVFLDSQATAANPVFTVFKDNE ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK LFAVKIPEQCT | NivG.798 |
| 447 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQQYTRSTDNQA VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK ISQSTASINENVNEKCKFTLPPLKIHECNISCPNPLPFREYRPQTEGV SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILNSTYWSGSLM MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG DTLYFPAVGFLVRTEFKYNDSNCPITKCQYSKPENCRLSMGIRPNS HYILRSGLLKYNLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF YQASFSWDTMIKFGDVLTVNPLVVNWRNNTVISRPGQSQCPRFNT CPAICAEGVYNDAFLIDRINWISAGVFLDSQATAANPVFTVFKDNE ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK LFAVKIPEQCT | NivG.799 |
| 448 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQQYTRSTDNQA VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK ISQSTASINENVNEKCKFTLPPLKIHECNISCPNPLPFREYRPQTEGV SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILNSTYWSGSLM MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG DTLYFPAVGFLVRTEFKYNDSNCPITKCQYSKPENCRLSMGIRPNS HYILRSGLLKYNLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF YQASFSWDTMIKFGDVLTVNPLVVNWRQNTVISRPGQSQCPRFNT CPAICAEGVYNDAFLIDRINWISAGVFLDSNATAANPVFTVFKDNE | NivG.800 |

| | | |
|---|---|---|
| | ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK<br>LFAVKIPEQCT | |
| 449 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQQYTRSTDNQA<br>VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK<br>ISQSTASINENVNEKCKFTLPPLKIHECNISCPNPLPFREYRPQTEGV<br>SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD<br>EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW<br>TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILNSTYWSGSLM<br>MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG<br>DTLYFPAVGFLVRTEFKYNDSNCPITKCQYSKPENCRLSMGIRPNS<br>HYILRSGLLKYNLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF<br>YQASFSWDTMIKFGDVLTVNPLVVNWRQNTVISRPGQSQCPRFNT<br>CPAICAEGVYNDAFLIDRINWISAGVFLDSQATAANPVFTVFKDNE<br>ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK<br>LFAVKIPEQCT | NivG.801 |
| 450 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQQYTRSTDNQA<br>VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK<br>ISQSTASINENVNEKCKFTLPPLKIHECNISCPNPLPFREYRPQTEGV<br>SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD<br>EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW<br>TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILNSTYWSGSLM<br>MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG<br>DTLYFPAVGFLVRTEFKYNDSNCPITKCQYSKPENCRLSMGIRPNS<br>HYILRSGLLKYQLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF<br>YQASFSWDTMIKFGDVLTVNPLVVNWRNNTVISRPGQSQCPRFNT<br>CPAICAEGVYNDAFLIDRINWISAGVFLDSNATAANPVFTVFKDNE<br>ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK<br>LFAVKIPEQCT | NivG.802 |
| 451 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQQYTRSTDNQA<br>VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK<br>ISQSTASINENVNEKCKFTLPPLKIHECNISCPNPLPFREYRPQTEGV<br>SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD<br>EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW<br>TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILNSTYWSGSLM<br>MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG<br>DTLYFPAVGFLVRTEFKYNDSNCPITKCQYSKPENCRLSMGIRPNS<br>HYILRSGLLKYQLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF<br>YQASFSWDTMIKFGDVLTVNPLVVNWRNNTVISRPGQSQCPRFNT<br>CPAICAEGVYNDAFLIDRINWISAGVFLDSQATAANPVFTVFKDNE<br>ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK<br>LFAVKIPEQCT | NivG.803 |
| 452 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQQYTRSTDNQA<br>VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK<br>ISQSTASINENVNEKCKFTLPPLKIHECNISCPNPLPFREYRPQTEGV<br>SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD<br>EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW<br>TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILNSTYWSGSLM<br>MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG<br>DTLYFPAVGFLVRTEFKYNDSNCPITKCQYSKPENCRLSMGIRPNS<br>HYILRSGLLKYQLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF<br>YQASFSWDTMIKFGDVLTVNPLVVNWRQNTVISRPGQSQCPRFNT<br>CPAICAEGVYNDAFLIDRINWISAGVFLDSNATAANPVFTVFKDNE | NivG.804 |

| | | |
|---|---|---|
| | ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK LFAVKIPEQCT | |
| 453 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQQYTRSTDNQA VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK ISQSTASINENVNEKCKFTLPPLKIHECNISCPNPLPFREYRPQTEGV SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILNSTYWSGSLM MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG DTLYFPAVGFLVRTEFKYNDSNCPITKCQYSKPENCRLSMGIRPNS HYILRSGLLKYQLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF YQASFSWDTMIKFGDVLTVNPLVVNWRQNTVISRPGQSQCPRFNT CPAICAEGVYNDAFLIDRINWISAGVFLDSQATAANPVFTVFKDNE ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK LFAVKIPEQCT | NivG.805 |
| 454 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQQYTRSTDNQA VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK ISQSTASINENVNEKCKFTLPPLKIHECNISCPNPLPFREYRPQTEGV SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILNSTYWSGSLM MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG DTLYFPAVGFLVRTEFKYQDSNCPITKCQYSKPENCRLSMGIRPNS HYILRSGLLKYNLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF YQASFSWDTMIKFGDVLTVNPLVVNWRNNTVISRPGQSQCPRFNT CPAICAEGVYNDAFLIDRINWISAGVFLDSNATAANPVFTVFKDNE ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK LFAVKIPEQCT | NivG.806 |
| 455 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQQYTRSTDNQA VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK ISQSTASINENVNEKCKFTLPPLKIHECNISCPNPLPFREYRPQTEGV SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILNSTYWSGSLM MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG DTLYFPAVGFLVRTEFKYQDSNCPITKCQYSKPENCRLSMGIRPNS HYILRSGLLKYNLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF YQASFSWDTMIKFGDVLTVNPLVVNWRNNTVISRPGQSQCPRFNT CPAICAEGVYNDAFLIDRINWISAGVFLDSQATAANPVFTVFKDNE ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK LFAVKIPEQCT | NivG.807 |
| 456 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQQYTRSTDNQA VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK ISQSTASINENVNEKCKFTLPPLKIHECNISCPNPLPFREYRPQTEGV SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILNSTYWSGSLM MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG DTLYFPAVGFLVRTEFKYQDSNCPITKCQYSKPENCRLSMGIRPNS HYILRSGLLKYNLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF YQASFSWDTMIKFGDVLTVNPLVVNWRQNTVISRPGQSQCPRFNT CPAICAEGVYNDAFLIDRINWISAGVFLDSNATAANPVFTVFKDNE | NivG.808 |

| | | |
|---|---|---|
| | ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK<br>LFAVKIPEQCT | |
| 457 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQQYTRSTDNQA<br>VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK<br>ISQSTASINENVNEKCKFTLPPLKIHECNISCPNPLPFREYRPQTEGV<br>SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD<br>EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW<br>TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILNSTYWSGSLM<br>MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG<br>DTLYFPAVGFLVRTEFKYQDSNCPITKCQYSKPENCRLSMGIRPNS<br>HYILRSGLLKYNLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF<br>YQASFSWDTMIKFGDVLTVNPLVVNWRQNTVISRPGQSQCPRFNT<br>CPAICAEGVYNDAFLIDRINWISAGVFLDSQATAANPVFTVFKDNE<br>ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK<br>LFAVKIPEQCT | NivG.809 |
| 458 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQQYTRSTDNQA<br>VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK<br>ISQSTASINENVNEKCKFTLPPLKIHECNISCPNPLPFREYRPQTEGV<br>SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD<br>EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW<br>TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILNSTYWSGSLM<br>MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG<br>DTLYFPAVGFLVRTEFKYQDSNCPITKCQYSKPENCRLSMGIRPNS<br>HYILRSGLLKYQLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF<br>YQASFSWDTMIKFGDVLTVNPLVVNWRNNTVISRPGQSQCPRFNT<br>CPAICAEGVYNDAFLIDRINWISAGVFLDSNATAANPVFTVFKDNE<br>ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK<br>LFAVKIPEQCT | NivG.810 |
| 459 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQQYTRSTDNQA<br>VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK<br>ISQSTASINENVNEKCKFTLPPLKIHECNISCPNPLPFREYRPQTEGV<br>SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD<br>EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW<br>TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILNSTYWSGSLM<br>MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG<br>DTLYFPAVGFLVRTEFKYQDSNCPITKCQYSKPENCRLSMGIRPNS<br>HYILRSGLLKYQLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF<br>YQASFSWDTMIKFGDVLTVNPLVVNWRNNTVISRPGQSQCPRFNT<br>CPAICAEGVYNDAFLIDRINWISAGVFLDSQATAANPVFTVFKDNE<br>ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK<br>LFAVKIPEQCT | NivG.811 |
| 460 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQQYTRSTDNQA<br>VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK<br>ISQSTASINENVNEKCKFTLPPLKIHECNISCPNPLPFREYRPQTEGV<br>SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD<br>EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW<br>TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILNSTYWSGSLM<br>MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG<br>DTLYFPAVGFLVRTEFKYQDSNCPITKCQYSKPENCRLSMGIRPNS<br>HYILRSGLLKYQLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF<br>YQASFSWDTMIKFGDVLTVNPLVVNWRQNTVISRPGQSQCPRFNT<br>CPAICAEGVYNDAFLIDRINWISAGVFLDSNATAANPVFTVFKDNE | NivG.812 |

| | | |
|---|---|---|
| | ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK LFAVKIPEQCT | |
| 461 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQQYTRSTDNQA VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK ISQSTASINENVNEKCKFTLPPLKIHECNISCPNPLPFREYRPQTEGV SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILNSTYWSGSLM MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG DTLYFPAVGFLVRTEFKYQDSNCPITKCQYSKPENCRLSMGIRPNS HYILRSGLLKYQLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF YQASFSWDTMIKFGDVLTVNPLVVNWRQNTVISRPGQSQCPRFNT CPAICAEGVYNDAFLIDRINWISAGVFLDSQATAANPVFTVFKDNE ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK LFAVKIPEQCT | NivG.813 |
| 462 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQQYTRSTDNQA VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK ISQSTASINENVNEKCKFTLPPLKIHECNISCPNPLPFREYRPQTEGV SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILQSTYWSGSLM MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG DTLYFPAVGFLVRTEFKYNDSNCPITKCQYSKPENCRLSMGIRPNS HYILRSGLLKYNLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF YQASFSWDTMIKFGDVLTVNPLVVNWRNNTVISRPGQSQCPRFNT CPAICAEGVYNDAFLIDRINWISAGVFLDSNATAANPVFTVFKDNE ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK LFAVKIPEQCT | NivG.814 |
| 463 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQQYTRSTDNQA VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK ISQSTASINENVNEKCKFTLPPLKIHECNISCPNPLPFREYRPQTEGV SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILQSTYWSGSLM MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG DTLYFPAVGFLVRTEFKYNDSNCPITKCQYSKPENCRLSMGIRPNS HYILRSGLLKYNLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF YQASFSWDTMIKFGDVLTVNPLVVNWRNNTVISRPGQSQCPRFNT CPAICAEGVYNDAFLIDRINWISAGVFLDSQATAANPVFTVFKDNE ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK LFAVKIPEQCT | NivG.815 |
| 464 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQQYTRSTDNQA VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK ISQSTASINENVNEKCKFTLPPLKIHECNISCPNPLPFREYRPQTEGV SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILQSTYWSGSLM MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG DTLYFPAVGFLVRTEFKYNDSNCPITKCQYSKPENCRLSMGIRPNS HYILRSGLLKYNLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF YQASFSWDTMIKFGDVLTVNPLVVNWRQNTVISRPGQSQCPRFNT CPAICAEGVYNDAFLIDRINWISAGVFLDSNATAANPVFTVFKDNE | NivG.816 |

| | | |
|---|---|---|
| | ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK<br>LFAVKIPEQCT | |
| 465 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQQYTRSTDNQA<br>VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK<br>ISQSTASINENVNEKCKFTLPPLKIHECNISCPNPLPFREYRPQTEGV<br>SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD<br>EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW<br>TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILQSTYWSGSLM<br>MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG<br>DTLYFPAVGFLVRTEFKYNDSNCPITKCQYSKPENCRLSMGIRPNS<br>HYILRSGLLKYNLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF<br>YQASFSWDTMIKFGDVLTVNPLVVNWRQNTVISRPGQSQCPRFNT<br>CPAICAEGVYNDAFLIDRINWISAGVFLDSQATAANPVFTVFKDNE<br>ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK<br>LFAVKIPEQCT | NivG.817 |
| 466 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQQYTRSTDNQA<br>VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK<br>ISQSTASINENVNEKCKFTLPPLKIHECNISCPNPLPFREYRPQTEGV<br>SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD<br>EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW<br>TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILQSTYWSGSLM<br>MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG<br>DTLYFPAVGFLVRTEFKYNDSNCPITKCQYSKPENCRLSMGIRPNS<br>HYILRSGLLKYQLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF<br>YQASFSWDTMIKFGDVLTVNPLVVNWRNNTVISRPGQSQCPRFNT<br>CPAICAEGVYNDAFLIDRINWISAGVFLDSNATAANPVFTVFKDNE<br>ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK<br>LFAVKIPEQCT | NivG.818 |
| 467 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQQYTRSTDNQA<br>VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK<br>ISQSTASINENVNEKCKFTLPPLKIHECNISCPNPLPFREYRPQTEGV<br>SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD<br>EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW<br>TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILQSTYWSGSLM<br>MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG<br>DTLYFPAVGFLVRTEFKYNDSNCPITKCQYSKPENCRLSMGIRPNS<br>HYILRSGLLKYQLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF<br>YQASFSWDTMIKFGDVLTVNPLVVNWRNNTVISRPGQSQCPRFNT<br>CPAICAEGVYNDAFLIDRINWISAGVFLDSQATAANPVFTVFKDNE<br>ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK<br>LFAVKIPEQCT | NivG.819 |
| 468 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQQYTRSTDNQA<br>VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK<br>ISQSTASINENVNEKCKFTLPPLKIHECNISCPNPLPFREYRPQTEGV<br>SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD<br>EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW<br>TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILQSTYWSGSLM<br>MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG<br>DTLYFPAVGFLVRTEFKYNDSNCPITKCQYSKPENCRLSMGIRPNS<br>HYILRSGLLKYQLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF<br>YQASFSWDTMIKFGDVLTVNPLVVNWRQNTVISRPGQSQCPRFNT<br>CPAICAEGVYNDAFLIDRINWISAGVFLDSNATAANPVFTVFKDNE | NivG.820 |

| | | |
|---|---|---|
| | ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK<br>LFAVKIPEQCT | |
| 469 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQQYTRSTDNQA<br>VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK<br>ISQSTASINENVNEKCKFTLPPLKIHECNISCPNPLPFREYRPQTEGV<br>SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD<br>EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW<br>TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILQSTYWSGSLM<br>MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG<br>DTLYFPAVGFLVRTEFKYNDSNCPITKCQYSKPENCRLSMGIRPNS<br>HYILRSGLLKYQLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF<br>YQASFSWDTMIKFGDVLTVNPLVVNWRQNTVISRPGQSQCPRFNT<br>CPAICAEGVYNDAFLIDRINWISAGVFLDSQATAANPVFTVFKDNE<br>ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK<br>LFAVKIPEQCT | NivG.821 |
| 470 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQQYTRSTDNQA<br>VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK<br>ISQSTASINENVNEKCKFTLPPLKIHECNISCPNPLPFREYRPQTEGV<br>SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD<br>EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW<br>TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILQSTYWSGSLM<br>MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG<br>DTLYFPAVGFLVRTEFKYQDSNCPITKCQYSKPENCRLSMGIRPNS<br>HYILRSGLLKYNLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF<br>YQASFSWDTMIKFGDVLTVNPLVVNWRNNTVISRPGQSQCPRFNT<br>CPAICAEGVYNDAFLIDRINWISAGVFLDSNATAANPVFTVFKDNE<br>ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK<br>LFAVKIPEQCT | NivG.822 |
| 471 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQQYTRSTDNQA<br>VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK<br>ISQSTASINENVNEKCKFTLPPLKIHECNISCPNPLPFREYRPQTEGV<br>SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD<br>EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW<br>TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILQSTYWSGSLM<br>MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG<br>DTLYFPAVGFLVRTEFKYQDSNCPITKCQYSKPENCRLSMGIRPNS<br>HYILRSGLLKYNLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF<br>YQASFSWDTMIKFGDVLTVNPLVVNWRNNTVISRPGQSQCPRFNT<br>CPAICAEGVYNDAFLIDRINWISAGVFLDSQATAANPVFTVFKDNE<br>ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK<br>LFAVKIPEQCT | NivG.823 |
| 472 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQQYTRSTDNQA<br>VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK<br>ISQSTASINENVNEKCKFTLPPLKIHECNISCPNPLPFREYRPQTEGV<br>SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD<br>EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW<br>TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILQSTYWSGSLM<br>MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG<br>DTLYFPAVGFLVRTEFKYQDSNCPITKCQYSKPENCRLSMGIRPNS<br>HYILRSGLLKYNLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF<br>YQASFSWDTMIKFGDVLTVNPLVVNWRQNTVISRPGQSQCPRFNT<br>CPAICAEGVYNDAFLIDRINWISAGVFLDSNATAANPVFTVFKDNE | NivG.824 |

| | | |
|---|---|---|
| | ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK<br>LFAVKIPEQCT | |
| 473 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQQYTRSTDNQA<br>VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK<br>ISQSTASINENVNEKCKFTLPPLKIHECNISCPNPLPFREYRPQTEGV<br>SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD<br>EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW<br>TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILQSTYWSGSLM<br>MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG<br>DTLYFPAVGFLVRTEFKYQDSNCPITKCQYSKPENCRLSMGIRPNS<br>HYILRSGLLKYNLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF<br>YQASFSWDTMIKFGDVLTVNPLVVNWRQNTVISRPGQSQCPRFNT<br>CPAICAEGVYNDAFLIDRINWISAGVFLDSQATAANPVFTVFKDNE<br>ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK<br>LFAVKIPEQCT | NivG.825 |
| 474 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQQYTRSTDNQA<br>VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK<br>ISQSTASINENVNEKCKFTLPPLKIHECNISCPNPLPFREYRPQTEGV<br>SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD<br>EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW<br>TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILQSTYWSGSLM<br>MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG<br>DTLYFPAVGFLVRTEFKYQDSNCPITKCQYSKPENCRLSMGIRPNS<br>HYILRSGLLKYQLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF<br>YQASFSWDTMIKFGDVLTVNPLVVNWRNNTVISRPGQSQCPRFNT<br>CPAICAEGVYNDAFLIDRINWISAGVFLDSNATAANPVFTVFKDNE<br>ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK<br>LFAVKIPEQCT | NivG.826 |
| 475 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQQYTRSTDNQA<br>VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK<br>ISQSTASINENVNEKCKFTLPPLKIHECNISCPNPLPFREYRPQTEGV<br>SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD<br>EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW<br>TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILQSTYWSGSLM<br>MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG<br>DTLYFPAVGFLVRTEFKYQDSNCPITKCQYSKPENCRLSMGIRPNS<br>HYILRSGLLKYQLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF<br>YQASFSWDTMIKFGDVLTVNPLVVNWRNNTVISRPGQSQCPRFNT<br>CPAICAEGVYNDAFLIDRINWISAGVFLDSQATAANPVFTVFKDNE<br>ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK<br>LFAVKIPEQCT | NivG.827 |
| 476 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQQYTRSTDNQA<br>VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK<br>ISQSTASINENVNEKCKFTLPPLKIHECNISCPNPLPFREYRPQTEGV<br>SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD<br>EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW<br>TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILQSTYWSGSLM<br>MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG<br>DTLYFPAVGFLVRTEFKYQDSNCPITKCQYSKPENCRLSMGIRPNS<br>HYILRSGLLKYQLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF<br>YQASFSWDTMIKFGDVLTVNPLVVNWRQNTVISRPGQSQCPRFNT<br>CPAICAEGVYNDAFLIDRINWISAGVFLDSNATAANPVFTVFKDNE | NivG.828 |

| | | |
|---|---|---|
| | ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK LFAVKIPEQCT | |
| 477 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQQYTRSTDNQA VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK ISQSTASINENVNEKCKFTLPPLKIHECNISCPNPLPFREYRPQTEGV SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILQSTYWSGSLM MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG DTLYFPAVGFLVRTEFKYQDSNCPITKCQYSKPENCRLSMGIRPNS HYILRSGLLKYQLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF YQASFSWDTMIKFGDVLTVNPLVVNWRQNTVISRPGQSQCPRFNT CPAICAEGVYNDAFLIDRINWISAGVFLDSQATAANPVFTVFKDNE ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK LFAVKIPEQCT | NivG.829 |
| 478 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQQYTRSTDNQA VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK ISQSTASINENVNEKCKFTLPPLKIHECQISCPNPLPFREYRPQTEGV SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILNSTYWSGSLM MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG DTLYFPAVGFLVRTEFKYNDSNCPITKCQYSKPENCRLSMGIRPNS HYILRSGLLKYNLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF YQASFSWDTMIKFGDVLTVNPLVVNWRNNTVISRPGQSQCPRFNT CPAICAEGVYNDAFLIDRINWISAGVFLDSNATAANPVFTVFKDNE ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK LFAVKIPEQCT | NivG.830 |
| 479 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQQYTRSTDNQA VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK ISQSTASINENVNEKCKFTLPPLKIHECQISCPNPLPFREYRPQTEGV SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILNSTYWSGSLM MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG DTLYFPAVGFLVRTEFKYNDSNCPITKCQYSKPENCRLSMGIRPNS HYILRSGLLKYNLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF YQASFSWDTMIKFGDVLTVNPLVVNWRNNTVISRPGQSQCPRFNT CPAICAEGVYNDAFLIDRINWISAGVFLDSQATAANPVFTVFKDNE ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK LFAVKIPEQCT | NivG.831 |
| 480 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQQYTRSTDNQA VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK ISQSTASINENVNEKCKFTLPPLKIHECQISCPNPLPFREYRPQTEGV SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILNSTYWSGSLM MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG DTLYFPAVGFLVRTEFKYNDSNCPITKCQYSKPENCRLSMGIRPNS HYILRSGLLKYNLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF YQASFSWDTMIKFGDVLTVNPLVVNWRQNTVISRPGQSQCPRFNT CPAICAEGVYNDAFLIDRINWISAGVFLDSNATAANPVFTVFKDNE | NivG.832 |

| | | |
|---|---|---|
| | ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK<br>LFAVKIPEQCT | |
| 481 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQQYTRSTDNQA<br>VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK<br>ISQSTASINENVNEKCKFTLPPLKIHECQISCPNPLPFREYRPQTEGV<br>SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD<br>EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW<br>TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILNSTYWSGSLM<br>MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG<br>DTLYFPAVGFLVRTEFKYNDSNCPITKCQYSKPENCRLSMGIRPNS<br>HYILRSGLLKYNLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF<br>YQASFSWDTMIKFGDVLTVNPLVVNWRQNTVISRPGQSQCPRFNT<br>CPAICAEGVYNDAFLIDRINWISAGVFLDSQATAANPVFTVFKDNE<br>ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK<br>LFAVKIPEQCT | NivG.833 |
| 482 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQQYTRSTDNQA<br>VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK<br>ISQSTASINENVNEKCKFTLPPLKIHECQISCPNPLPFREYRPQTEGV<br>SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD<br>EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW<br>TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILNSTYWSGSLM<br>MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG<br>DTLYFPAVGFLVRTEFKYNDSNCPITKCQYSKPENCRLSMGIRPNS<br>HYILRSGLLKYQLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF<br>YQASFSWDTMIKFGDVLTVNPLVVNWRNNTVISRPGQSQCPRFNT<br>CPAICAEGVYNDAFLIDRINWISAGVFLDSNATAANPVFTVFKDNE<br>ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK<br>LFAVKIPEQCT | NivG.834 |
| 483 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQQYTRSTDNQA<br>VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK<br>ISQSTASINENVNEKCKFTLPPLKIHECQISCPNPLPFREYRPQTEGV<br>SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD<br>EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW<br>TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILNSTYWSGSLM<br>MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG<br>DTLYFPAVGFLVRTEFKYNDSNCPITKCQYSKPENCRLSMGIRPNS<br>HYILRSGLLKYQLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF<br>YQASFSWDTMIKFGDVLTVNPLVVNWRNNTVISRPGQSQCPRFNT<br>CPAICAEGVYNDAFLIDRINWISAGVFLDSQATAANPVFTVFKDNE<br>ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK<br>LFAVKIPEQCT | NivG.835 |
| 484 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQQYTRSTDNQA<br>VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK<br>ISQSTASINENVNEKCKFTLPPLKIHECQISCPNPLPFREYRPQTEGV<br>SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD<br>EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW<br>TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILNSTYWSGSLM<br>MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG<br>DTLYFPAVGFLVRTEFKYNDSNCPITKCQYSKPENCRLSMGIRPNS<br>HYILRSGLLKYQLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF<br>YQASFSWDTMIKFGDVLTVNPLVVNWRQNTVISRPGQSQCPRFNT<br>CPAICAEGVYNDAFLIDRINWISAGVFLDSNATAANPVFTVFKDNE | NivG.836 |

| | | |
|---|---|---|
| | ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK<br>LFAVKIPEQCT | |
| 485 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQQYTRSTDNQA<br>VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK<br>ISQSTASINENVNEKCKFTLPPLKIHECQISCPNPLPFREYRPQTEGV<br>SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD<br>EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW<br>TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILNSTYWSGSLM<br>MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG<br>DTLYFPAVGFLVRTEFKYNDSNCPITKCQYSKPENCRLSMGIRPNS<br>HYILRSGLLKYQLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF<br>YQASFSWDTMIKFGDVLTVNPLVVNWRQNTVISRPGQSQCPRFNT<br>CPAICAEGVYNDAFLIDRINWISAGVFLDSQATAANPVFTVFKDNE<br>ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK<br>LFAVKIPEQCT | NivG.837 |
| 486 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQQYTRSTDNQA<br>VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK<br>ISQSTASINENVNEKCKFTLPPLKIHECQISCPNPLPFREYRPQTEGV<br>SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD<br>EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW<br>TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILNSTYWSGSLM<br>MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG<br>DTLYFPAVGFLVRTEFKYQDSNCPITKCQYSKPENCRLSMGIRPNS<br>HYILRSGLLKYNLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF<br>YQASFSWDTMIKFGDVLTVNPLVVNWRNNTVISRPGQSQCPRFNT<br>CPAICAEGVYNDAFLIDRINWISAGVFLDSNATAANPVFTVFKDNE<br>ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK<br>LFAVKIPEQCT | NivG.838 |
| 487 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQQYTRSTDNQA<br>VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK<br>ISQSTASINENVNEKCKFTLPPLKIHECQISCPNPLPFREYRPQTEGV<br>SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD<br>EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW<br>TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILNSTYWSGSLM<br>MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG<br>DTLYFPAVGFLVRTEFKYQDSNCPITKCQYSKPENCRLSMGIRPNS<br>HYILRSGLLKYNLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF<br>YQASFSWDTMIKFGDVLTVNPLVVNWRNNTVISRPGQSQCPRFNT<br>CPAICAEGVYNDAFLIDRINWISAGVFLDSQATAANPVFTVFKDNE<br>ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK<br>LFAVKIPEQCT | NivG.839 |
| 488 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQQYTRSTDNQA<br>VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK<br>ISQSTASINENVNEKCKFTLPPLKIHECQISCPNPLPFREYRPQTEGV<br>SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD<br>EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW<br>TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILNSTYWSGSLM<br>MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG<br>DTLYFPAVGFLVRTEFKYQDSNCPITKCQYSKPENCRLSMGIRPNS<br>HYILRSGLLKYNLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF<br>YQASFSWDTMIKFGDVLTVNPLVVNWRQNTVISRPGQSQCPRFNT<br>CPAICAEGVYNDAFLIDRINWISAGVFLDSNATAANPVFTVFKDNE | NivG.840 |

| | | |
|---|---|---|
| | ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK LFAVKIPEQCT | |
| 489 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQQYTRSTDNQA VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK ISQSTASINENVNEKCKFTLPPLKIHECQISCPNPLPFREYRPQTEGV SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILNSTYWSGSLM MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG DTLYFPAVGFLVRTEFKYQDSNCPITKCQYSKPENCRLSMGIRPNS HYILRSGLLKYNLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF YQASFSWDTMIKFGDVLTVNPLVVNWRQNTVISRPGQSQCPRFNT CPAICAEGVYNDAFLIDRINWISAGVFLDSQATAANPVFTVFKDNE ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK LFAVKIPEQCT | NivG.841 |
| 490 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQQYTRSTDNQA VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK ISQSTASINENVNEKCKFTLPPLKIHECQISCPNPLPFREYRPQTEGV SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILNSTYWSGSLM MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG DTLYFPAVGFLVRTEFKYQDSNCPITKCQYSKPENCRLSMGIRPNS HYILRSGLLKYQLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF YQASFSWDTMIKFGDVLTVNPLVVNWRNNTVISRPGQSQCPRFNT CPAICAEGVYNDAFLIDRINWISAGVFLDSNATAANPVFTVFKDNE ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK LFAVKIPEQCT | NivG.842 |
| 491 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQQYTRSTDNQA VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK ISQSTASINENVNEKCKFTLPPLKIHECQISCPNPLPFREYRPQTEGV SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILNSTYWSGSLM MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG DTLYFPAVGFLVRTEFKYQDSNCPITKCQYSKPENCRLSMGIRPNS HYILRSGLLKYQLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF YQASFSWDTMIKFGDVLTVNPLVVNWRNNTVISRPGQSQCPRFNT CPAICAEGVYNDAFLIDRINWISAGVFLDSQATAANPVFTVFKDNE ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK LFAVKIPEQCT | NivG.843 |
| 492 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQQYTRSTDNQA VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK ISQSTASINENVNEKCKFTLPPLKIHECQISCPNPLPFREYRPQTEGV SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILNSTYWSGSLM MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG DTLYFPAVGFLVRTEFKYQDSNCPITKCQYSKPENCRLSMGIRPNS HYILRSGLLKYQLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF YQASFSWDTMIKFGDVLTVNPLVVNWRQNTVISRPGQSQCPRFNT CPAICAEGVYNDAFLIDRINWISAGVFLDSNATAANPVFTVFKDNE | NivG.844 |

| | | |
|---|---|---|
| | ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK<br>LFAVKIPEQCT | |
| 493 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQQYTRSTDNQA<br>VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK<br>ISQSTASINENVNEKCKFTLPPLKIHECQISCPNPLPFREYRPQTEGV<br>SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD<br>EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW<br>TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILNSTYWSGSLM<br>MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG<br>DTLYFPAVGFLVRTEFKYQDSNCPITKCQYSKPENCRLSMGIRPNS<br>HYILRSGLLKYQLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF<br>YQASFSWDTMIKFGDVLTVNPLVVNWRQNTVISRPGQSQCPRFNT<br>CPAICAEGVYNDAFLIDRINWISAGVFLDSQATAANPVFTVFKDNE<br>ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK<br>LFAVKIPEQCT | NivG.845 |
| 494 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQQYTRSTDNQA<br>VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK<br>ISQSTASINENVNEKCKFTLPPLKIHECQISCPNPLPFREYRPQTEGV<br>SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD<br>EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW<br>TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILQSTYWSGSLM<br>MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG<br>DTLYFPAVGFLVRTEFKYNDSNCPITKCQYSKPENCRLSMGIRPNS<br>HYILRSGLLKYNLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF<br>YQASFSWDTMIKFGDVLTVNPLVVNWRNNTVISRPGQSQCPRFNT<br>CPAICAEGVYNDAFLIDRINWISAGVFLDSNATAANPVFTVFKDNE<br>ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK<br>LFAVKIPEQCT | NivG.846 |
| 495 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQQYTRSTDNQA<br>VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK<br>ISQSTASINENVNEKCKFTLPPLKIHECQISCPNPLPFREYRPQTEGV<br>SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD<br>EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW<br>TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILQSTYWSGSLM<br>MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG<br>DTLYFPAVGFLVRTEFKYNDSNCPITKCQYSKPENCRLSMGIRPNS<br>HYILRSGLLKYNLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF<br>YQASFSWDTMIKFGDVLTVNPLVVNWRNNTVISRPGQSQCPRFNT<br>CPAICAEGVYNDAFLIDRINWISAGVFLDSQATAANPVFTVFKDNE<br>ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK<br>LFAVKIPEQCT | NivG.847 |
| 496 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQQYTRSTDNQA<br>VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK<br>ISQSTASINENVNEKCKFTLPPLKIHECQISCPNPLPFREYRPQTEGV<br>SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD<br>EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW<br>TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILQSTYWSGSLM<br>MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG<br>DTLYFPAVGFLVRTEFKYNDSNCPITKCQYSKPENCRLSMGIRPNS<br>HYILRSGLLKYNLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF<br>YQASFSWDTMIKFGDVLTVNPLVVNWRQNTVISRPGQSQCPRFNT<br>CPAICAEGVYNDAFLIDRINWISAGVFLDSNATAANPVFTVFKDNE | NivG.848 |

| | | |
|---|---|---|
| | ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK<br>LFAVKIPEQCT | |
| 497 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQQYTRSTDNQA<br>VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK<br>ISQSTASINENVNEKCKFTLPPLKIHECQISCPNPLPFREYRPQTEGV<br>SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD<br>EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW<br>TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILQSTYWSGSLM<br>MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG<br>DTLYFPAVGFLVRTEFKYNDSNCPITKCQYSKPENCRLSMGIRPNS<br>HYILRSGLLKYNLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF<br>YQASFSWDTMIKFGDVLTVNPLVVNWRQNTVISRPGQSQCPRFNT<br>CPAICAEGVYNDAFLIDRINWISAGVFLDSQATAANPVFTVFKDNE<br>ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK<br>LFAVKIPEQCT | NivG.849 |
| 498 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQQYTRSTDNQA<br>VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK<br>ISQSTASINENVNEKCKFTLPPLKIHECQISCPNPLPFREYRPQTEGV<br>SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD<br>EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW<br>TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILQSTYWSGSLM<br>MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG<br>DTLYFPAVGFLVRTEFKYNDSNCPITKCQYSKPENCRLSMGIRPNS<br>HYILRSGLLKYQLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF<br>YQASFSWDTMIKFGDVLTVNPLVVNWRNNTVISRPGQSQCPRFNT<br>CPAICAEGVYNDAFLIDRINWISAGVFLDSNATAANPVFTVFKDNE<br>ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK<br>LFAVKIPEQCT | NivG.850 |
| 499 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQQYTRSTDNQA<br>VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK<br>ISQSTASINENVNEKCKFTLPPLKIHECQISCPNPLPFREYRPQTEGV<br>SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD<br>EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW<br>TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILQSTYWSGSLM<br>MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG<br>DTLYFPAVGFLVRTEFKYNDSNCPITKCQYSKPENCRLSMGIRPNS<br>HYILRSGLLKYQLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF<br>YQASFSWDTMIKFGDVLTVNPLVVNWRNNTVISRPGQSQCPRFNT<br>CPAICAEGVYNDAFLIDRINWISAGVFLDSQATAANPVFTVFKDNE<br>ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK<br>LFAVKIPEQCT | NivG.851 |
| 500 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQQYTRSTDNQA<br>VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK<br>ISQSTASINENVNEKCKFTLPPLKIHECQISCPNPLPFREYRPQTEGV<br>SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD<br>EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW<br>TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILQSTYWSGSLM<br>MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG<br>DTLYFPAVGFLVRTEFKYNDSNCPITKCQYSKPENCRLSMGIRPNS<br>HYILRSGLLKYQLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF<br>YQASFSWDTMIKFGDVLTVNPLVVNWRQNTVISRPGQSQCPRFNT<br>CPAICAEGVYNDAFLIDRINWISAGVFLDSNATAANPVFTVFKDNE | NivG.852 |

| | | |
|---|---|---|
| | ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK<br>LFAVKIPEQCT | |
| 501 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQQYTRSTDNQA<br>VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK<br>ISQSTASINENVNEKCKFTLPPLKIHECQISCPNPLPFREYRPQTEGV<br>SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD<br>EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW<br>TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILQSTYWSGSLM<br>MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG<br>DTLYFPAVGFLVRTEFKYNDSNCPITKCQYSKPENCRLSMGIRPNS<br>HYILRSGLLKYQLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF<br>YQASFSWDTMIKFGDVLTVNPLVVNWRQNTVISRPGQSQCPRFNT<br>CPAICAEGVYNDAFLIDRINWISAGVFLDSQATAANPVFTVFKDNE<br>ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK<br>LFAVKIPEQCT | NivG.853 |
| 502 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQQYTRSTDNQA<br>VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK<br>ISQSTASINENVNEKCKFTLPPLKIHECQISCPNPLPFREYRPQTEGV<br>SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD<br>EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW<br>TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILQSTYWSGSLM<br>MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG<br>DTLYFPAVGFLVRTEFKYQDSNCPITKCQYSKPENCRLSMGIRPNS<br>HYILRSGLLKYNLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF<br>YQASFSWDTMIKFGDVLTVNPLVVNWRNNTVISRPGQSQCPRFNT<br>CPAICAEGVYNDAFLIDRINWISAGVFLDSNATAANPVFTVFKDNE<br>ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK<br>LFAVKIPEQCT | NivG.854 |
| 503 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQQYTRSTDNQA<br>VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK<br>ISQSTASINENVNEKCKFTLPPLKIHECQISCPNPLPFREYRPQTEGV<br>SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD<br>EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW<br>TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILQSTYWSGSLM<br>MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG<br>DTLYFPAVGFLVRTEFKYQDSNCPITKCQYSKPENCRLSMGIRPNS<br>HYILRSGLLKYNLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF<br>YQASFSWDTMIKFGDVLTVNPLVVNWRNNTVISRPGQSQCPRFNT<br>CPAICAEGVYNDAFLIDRINWISAGVFLDSQATAANPVFTVFKDNE<br>ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK<br>LFAVKIPEQCT | NivG.855 |
| 504 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQQYTRSTDNQA<br>VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK<br>ISQSTASINENVNEKCKFTLPPLKIHECQISCPNPLPFREYRPQTEGV<br>SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD<br>EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW<br>TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILQSTYWSGSLM<br>MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG<br>DTLYFPAVGFLVRTEFKYQDSNCPITKCQYSKPENCRLSMGIRPNS<br>HYILRSGLLKYNLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF<br>YQASFSWDTMIKFGDVLTVNPLVVNWRQNTVISRPGQSQCPRFNT<br>CPAICAEGVYNDAFLIDRINWISAGVFLDSNATAANPVFTVFKDNE | NivG.856 |

| | | |
|---|---|---|
| | ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK LFAVKIPEQCT | |
| 505 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQQYTRSTDNQA VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK ISQSTASINENVNEKCKFTLPPLKIHECQISCPNPLPFREYRPQTEGV SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILQSTYWSGSLM MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG DTLYFPAVGFLVRTEFKYQDSNCPITKCQYSKPENCRLSMGIRPNS HYILRSGLLKYNLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF YQASFSWDTMIKFGDVLTVNPLVVNWRQNTVISRPGQSQCPRFNT CPAICAEGVYNDAFLIDRINWISAGVFLDSQATAANPVFTVFKDNE ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK LFAVKIPEQCT | NivG.857 |
| 506 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQQYTRSTDNQA VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK ISQSTASINENVNEKCKFTLPPLKIHECQISCPNPLPFREYRPQTEGV SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILQSTYWSGSLM MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG DTLYFPAVGFLVRTEFKYQDSNCPITKCQYSKPENCRLSMGIRPNS HYILRSGLLKYQLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF YQASFSWDTMIKFGDVLTVNPLVVNWRNNTVISRPGQSQCPRFNT CPAICAEGVYNDAFLIDRINWISAGVFLDSNATAANPVFTVFKDNE ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK LFAVKIPEQCT | NivG.858 |
| 507 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQQYTRSTDNQA VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK ISQSTASINENVNEKCKFTLPPLKIHECQISCPNPLPFREYRPQTEGV SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILQSTYWSGSLM MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG DTLYFPAVGFLVRTEFKYQDSNCPITKCQYSKPENCRLSMGIRPNS HYILRSGLLKYQLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF YQASFSWDTMIKFGDVLTVNPLVVNWRNNTVISRPGQSQCPRFNT CPAICAEGVYNDAFLIDRINWISAGVFLDSQATAANPVFTVFKDNE ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK LFAVKIPEQCT | NivG.859 |
| 508 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQQYTRSTDNQA VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK ISQSTASINENVNEKCKFTLPPLKIHECQISCPNPLPFREYRPQTEGV SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILQSTYWSGSLM MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG DTLYFPAVGFLVRTEFKYQDSNCPITKCQYSKPENCRLSMGIRPNS HYILRSGLLKYQLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF YQASFSWDTMIKFGDVLTVNPLVVNWRQNTVISRPGQSQCPRFNT CPAICAEGVYNDAFLIDRINWISAGVFLDSNATAANPVFTVFKDNE | NivG.860 |

| | | |
|---|---|---|
| | ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK LFAVKIPEQCT | |
| 509 | MSPQRDRINAFYKDNPHPKGSRIVINREHLMIDRPYVLLAVLFVMF LSLIGLLAIAGIRLHRAAIYTAEIHKSLSTNLDVTNSIEHQVKDVLTP LFKIIGDEVGLRTPQRFTDLVKFISDKIKFLNPDREYDFRDLTWCIN PPERIKLDYDQYCADVAAEELMNALVNSTLLETRTTNQFLAVSKG NCSGPTTIRGQFSNMSLSLLDLYLGRGYNVSSIVTMTSQGMYGGT YLVEKPNLSSKRSELSQLSMYRVFEVGVIRNPGLGAPVFHMTNYL EQPVSNDLSNCMVALGELKLAALCHGEDSITIPYQGSGKGVSFQL VKLGVWKSPTDMQSWVPLSTDD- PVIDRLYLSSHRGVIADNQAKWAVPTTRTDDKLRMETCFQQACK GKIQALCENPEWAPLKDNRIPSYGVLSVDLSLTVELKIKIASGFGPL ITHGSGMDLYKSNHNNVYWLTIPPMKNLALGVINTLEWIPRFKVS PYLFNVPIKEAGEDCHAPTYLPAEVDGDVKLSSNLVILPGQDLQYV LATYDTSRVEHAVVYYVYSPSRSFSYFYPFRLPIKGVPIELQVECFT WDQKLWCRHFCVLADSESGGHITHSGMEGMGVSCTVTREDGTN RR | Measles H protein |
| 510 | MLPYQDKVGAFYKDNARANSTKLSLVTEGHGGRRPPYLLFVLLIL LVGILALLAITGVRFHQVSTSNMEFSRLLKEDMEKSEAVHHQVIDV LTPLFKIIGDEIGLRLPQKLNEIKQFILQKTNFFNPNREFDFRDLHWC INPPSTVKVNFTNYCESIGIRKAIASAANPILLSALSGGRGDIFPPHR CSGATTSVGKVFPLSVSLSMSLISRTSEVINMLTAISDGVYGKTYLL VPDDIEREFDTREIRVFEIGFIKRWLNDMPLLQTTNYMVLPKNSKA KVCTIAVGELTLASLCVEESTVLLYHDSSGSQDGILVVTLGIFWAT PMDHIEEVIPVAHPSMKKIHITNHRGFIKDSIATWMVPALASEKQE EQKGCLESACQRKTYPMCNQASWEPFGGRQLPSYGRLTLPLDASV DLQLNISFTYGPVILNGDGMDYYESPLLNSGWLTIPPKDGTISGLIN KAGRGDQFTVLPHVLTFAPRESSGNCYLPIQTSQIRDRDVLIESNIV VLPTQSIRYVIATYDISRSDHAIVYYVYDPIRTISYTHPFRLTTKGRP DFLRIECFVWDDNLWCHQF | Canine distemper virus H protein |
| 511 | DIPATYEFTDGKHYITNEPIPPK | DogTag |
| 512 | KLGEIEFIKVDKTDKKPLRGAVFSLQKQHPDYPDIYGAIDQNGTYQ DVRTGEDGKLTFTNLSDGKYRLIENSEPPGYKPVQNKPIVSFRIVD GEVRDVTSIVPQ | DogCatcher |
| 513 | ATGAAGAAGATAAACGAGGGACTACTGGACAGTAAAATTCTCT CAGCCTTTAACACAGTAATCGCGCTCCTGGGCTCAATTGTGATT ATTGTTATGAATATAATGATTATACAGAACTACACCAGGAGCA CAGACAATCAGGCTGTGATCAAGGACGCACTTCAGGGCATCCA GCAACAGATCAAGGGGTTAGCGGATAAGATTGGAACTGAAATT GGGCCAAAGGTTTCACTGATTGATACCTCTAGCACGATCACAAT CCCCGCTAACATTGGCCTCCTGGGTTCCAAGATTAGCCAGTCTA CTGCAAGCATCAACGAAAACGTGAATGAGAAGTGTAAATTCAC CCTTCCCCCGCTGAAAATTCACGAATGCAACATCTCTTGTCCCA ATCCCCTGCCTTTCAGGGAGTATCGCCCACAAACCGAGGGGGT GTCTAATTTGGTCGGCTTACCTAATAACATCTGCTTGCAGAAAA CCTCCAATCAGATCCTGAAACCGAAGTTGATCTCCTACACTCTT CCCGTCGTCGGTCAAAGTGGGACTTGCATCACTGACCCTCTGCT GGCCATGGACGAGGGTTACTTCGCCTACAGTCACCTGGAGCGC ATCGGCTCCTGCAGCCGCGGGGTATCGAAACAGCGAATCATAG GAGTGGGAGAGGTACTCGATAGAGGTGACGAAGTGCCTAGCCT CTTTATGACCAACGTGTGGACACCACCAAACCCCAACACTGTGT ACCACTGTTCTGCAGTGTACAACAATGAGTTCTACTATGTGTTA TGCGCCGTGTCCACCGTGGGAGATCCCATCCTTAATAGCACATA | NivG388_SpyTag_391_ |

| | | |
|---|---|---|
| | CTGGTCCGGCTCTCTGATGATGACACGCCTGGCCGTAAAGCCCA<br>AAAGCAACGGGGGAGGGTACAACCAGCATCAGTTGGCTCTCAG<br>ATCCATTGAAAAGGGAAGATACGATAAAGTGATGCCATACGGC<br>CCATCTGGCATCAAGCAGGGGGATACCCTCTATTTTCCAGCAGT<br>GGGCTTTCTGGTGCGGACAGAATTCAAGTATAACGACTCAAAT<br>TGTCCAATTACCAAGTGCCAGGGAGGTGGAGGATCTGCCCACA<br>TCGTGATGGTCGACGCTTACAAGCCCACCAAAGGTGGAGGAGG<br>ATCTCCTGAGAATTGCAGGTTGAGTATGGGCATCCGTCCTAATA<br>GTCATTATATTCTCCGGAGTGGTCTGCTCAAATACAACCTGAGT<br>GATGGCGAGAACCCGAAAGTCGTCTTCATAGAGATATCAGATC<br>AACGGCTCTCAATTGGCAGCCCCTCCAAAATTTATGACAGTCTT<br>GGACAGCCTGTTTTTTATCAGGCTTCCTTTTCATGGGACACAAT<br>GATCAAGTTCGGCGATGTCCTGACGGTTAACCCGCTAGTTGTCA<br>ACTGGCGAAACAATACAGTGATCTCTAGACCTGGCCAAAGCCA<br>ATGTCCCCGTTTCAATACTTGTCCTGCGATCTGCGCTGAAGGTG<br>TGTATAATGACGCCTTCCTGATTGACCGGATAAATTGGATCTCG<br>GCAGGAGTCTTTCTAGATTCTAATGCCACCGCCGCTAATCCAGT<br>TTTTACGGTTTTCAAGGATAATGAAATTCTGTATAGAGCCCAGT<br>TGGCTTCCGAGGACACAAATGCCCAGAAGACCATAACGAACTG<br>TTTCCTGCTGAAAAACAAGATCTGGTGCATCAGCTTAGTCGAGA<br>TTTATGACACTGGAGATAACGTGATAAGGCCCAAGCTTTTTGCA<br>GTAAAAATCCCTGAACAGTGTACTTAG | |
| 514 | ATGAAGAAGATCAACGAGGGCCTGCTGGACAGCAAGATCCTGA<br>GCGCCTTCAACACCGTGATCGCCCTGCTGGGCAGCATCGTGATC<br>ATTGTGATGAACATCATGATCATCCAGAACTACACCAGAAGCA<br>CCGACAACCAGGCCGTGATCAAGGACGCTCTCCAGGGGATCCA<br>GCAGCAGATCAAGGGCCTGGCCGACAAGATCGGCACCGAGATC<br>GGCCCCAAGGTGTCCCTGATCGACACCAGCAGCACCATCACCA<br>TCCCCGCCAACATCGGCCTGCTGGGGTCCAAGATCAGCCAGAG<br>CACCGCCAGCATCAACGAGAACGTGAACGAGAAGTGCAAGTTC<br>ACCCTGCCCCCCCTGAAGATCCACGAGTGCAACATCAGCTGCC<br>CCAACCCCCTGCCCTTCCGGGAGTACCGGCCCCAGACCGAGGG<br>CGTGAGCAACCTGGTCGGCCTGCCCAACAACATCTGCCTGCAG<br>AAAACCAGCAACCAGATCCTGAAGCCCAAGCTCATTTCCTACA<br>CCCTGCCCGTGGTGGGCCAGAGCGGCACCTGCATCACCGACCC<br>CCTGCTGGCCATGGACGAGGGCTACTTCGCCTACAGCCACCTG<br>GAACGGATCGGCAGCTGCAGCAGGGGCGTGTCCAAGCAGCGG<br>ATCATCGGCGTGGGCGAGGTGCTGGACCGGGGCGACGAGGTGC<br>CCAGCCTGTTCATGACCAACGTGTGGACCCCCCCCAACCCCAAC<br>ACCGTGTACCACTGCAGCGCCGTGTACAACAACGAGTTCTACT<br>ACGTGCTGTGCGCCGTGAGCACCGTGGGCGACCCCATCCTGAA<br>CAGCACCTACTGGTCCGGCAGCCTGATGATGACCCGGCTGGCC<br>GTGAAGCCTAAGAGCAATGGCGGCGGATACAACCAGCACCAGC<br>TGGCCCTGCGGAGCATCGAGAAGGGCAGATACGACAAAGTGAT<br>GCCCTACGGCCCCAGCGGCATCAAGCAGGGCGACACACTGTAC<br>TTCCCCGCCGTGGGCTTCCTGGTCCGGACCGAGTTCAAGTACAA<br>CGACAGCAACTGCCCCATCACCAAGTGCCAGTACAGCAAGCCC<br>GAGAACTGCAGACTGAGCATGGGCATCCGGCCCAACAGCCACT<br>ACATCCTGCGGAGCGGCCTGCTGAAGTACAACCTGAGCGACGG<br>CGAGAACCCCAAAGTCGTCTTTATTGAGATCAGCGACCAGCGG<br>CTGTCCATCGGCAGCCCCAGCAAGATCTACGACAGCCTGGGCC<br>AGCCCGTGTTCTACCAGGCCAGCTTCAGCTGGGACACCATGATC<br>AAGTTCGGCGACGTGCTGACCGTGAACCCCCTGGTGGTGAACT | NivG.SpyTag_ |

| | | |
|---|---|---|
| | GGCGGAACAATACCGTGATCAGCAGACCCGGCCAGAGCCAGTG<br>CCCCCGGTTCAACACCTGCCCCGCGATCTGCGCGGAGGGCGTG<br>TACAACGACGCCTTCCTGATCGACCGGATCAACTGGATCTCTGC<br>CGGCGTGTTCCTGGACTCCAACGCGACCGCCGCGAATCCCGTGT<br>TCACCGTGTTTAAGGACAACGAGATCCTGTACCGGGCCCAGCT<br>GGCCAGCGAGGACACCAACGCCCAGAAAACCATCACCAACTGC<br>TTTCTGCTGAAGAACAAGATCTGGTGCATCAGCCTGGTGGAGA<br>TCTACGATACCGGCGACAACGTGATCAGGCCCAAGCTGTTCGC<br>CGTGAAGATCCCCGAGCAGTGCACCGGTGGTGGTGGATCTGGA<br>GGTGGTGGTAGCGCCCACATCGTGATGGTCGACGCTTACAAGC<br>CCACCAAATAG | |
| 515 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQNYTRSTDNQA<br>VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK<br>ISQSTASINENVNEKCKFTLPPLKIHECNISCPNPLPFREYRPQTEGV<br>SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD<br>EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW<br>TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILNSTYWSGSLM<br>MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG<br>DTLYFPAVGFLVRTEFKYNDSNCPITKCQGGGGSAHIVMVDAYKP<br>TKGGGGSPENCRLSMGIRPNSHYILRSGLLKYNLSDGENPKVVFIEI<br>SDQRLSIGSPSKIYDSLGQPVFYQASFSWDTMIKFGDVLTVNPLVV<br>NWRNNTVISRPGQSQCPRFNTCPAICAEGVYNDAFLIDRINWISAG<br>VFLDSNATAANPVFTVFKDNEILYRAQLASEDTNAQKTITNCFLLK<br>NKIWCISLVEIYDTGDNVIRPKLFAVKIPEQCT* | NivG388_SpyTag_391_ |
| 516 | MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQNYTRSTDNQA<br>VIKDALQGIQQQIKGLADKIGTEIGPKVSLIDTSSTITIPANIGLLGSK<br>ISQSTASINENVNEKCKFTLPPLKIHECNISCPNPLPFREYRPQTEGV<br>SNLVGLPNNICLQKTSNQILKPKLISYTLPVVGQSGTCITDPLLAMD<br>EGYFAYSHLERIGSCSRGVSKQRIIGVGEVLDRGDEVPSLFMTNVW<br>TPPNPNTVYHCSAVYNNEFYYVLCAVSTVGDPILNSTYWSGSLM<br>MTRLAVKPKSNGGGYNQHQLALRSIEKGRYDKVMPYGPSGIKQG<br>DTLYFPAVGFLVRTEFKYNDSNCPITKCQYSKPENCRLSMGIRPNS<br>HYILRSGLLKYNLSDGENPKVVFIEISDQRLSIGSPSKIYDSLGQPVF<br>YQASFSWDTMIKFGDVLTVNPLVVNWRNNTVISRPGQSQCPRFNT<br>CPAICAEGVYNDAFLIDRINWISAGVFLDSNATAANPVFTVFKDNE<br>ILYRAQLASEDTNAQKTITNCFLLKNKIWCISLVEIYDTGDNVIRPK<br>LFAVKIPEQCTGGGGSGGGGSAHIVMVDAYKPTK* | NivG.SpyTag_ |

## WHAT IS CLAIMED:

1. A lipid particle, comprising:

(a) a paramyxovirus envelope attachment protein attached to a targeting moiety via a universal adapter system, said system comprising a tag polypeptide component and a catcher polypeptide component, wherein the paramyxovirus envelope attachment protein is fused to one of the components, and the targeting moiety is fused to the other of the components; and

(b) at least one paramyxovirus fusion (F) protein;

wherein the protein in (a) and (b) are exposed on the outside of the lipid bilayer.

2. The lipid particle of claim 1, wherein the tag polypeptide component and catcher polypeptide component are covalently bound.

3. A lipid particle, comprising:

(a) a paramyxovirus envelope attachment protein fused to a tag polypeptide component or fused to a catcher polypeptide component of a universal adapter system; and

(b) at least one paramyxovirus fusion (F) protein;

wherein the protein in (a) and (b) are exposed on the outside of the lipid bilayer.

4. The lipid particle of any of claims 1-3, wherein the paramyxovirus envelope attachment protein is fused to the tag polypeptide component of the universal adapter system.

5. The lipid particle of any of claims 1-3, wherein the paramyxovirus envelope attachment protein is fused to the catcher polypeptide component of the universal adapter system.

6. The lipid particle of any of claims 1-5, wherein the paramyxovirus envelope attachment protein is fused to the component of the universal adapter system at the C terminus of the paramyxovirus envelope attachment protein.

7. The lipid particle of any of claims 1-5, wherein the paramyxovirus envelope attachment protein is fused to the component of the universal adapter system by insertion of the component within a peptide loop of the paramyxovirus envelope attachment protein.

8. The lipid particle of claim 7, wherein the peptide loop of the paramyxovirus envelope attachment protein is selected from the group consisting of:

(i) amino acid residues 194-197;

(ii) amino acid residues 208-213;

(iii) amino acid residues 241-243;

(iv) amino acid residues 242-243;

(v) amino acid residues 388-391;

(vi) amino acid residues 402-405;

(vii) amino acid residues 489-491; and

(viii) amino acid residues 495-497;

optionally wherein the peptide loop is 388-391, 489-491, or 495-497.

9. The lipid particle of claim 5-8, wherein the fusion is direct or indirect via flexible linker, optionally a peptide linker.

10. The lipid particle of claim 9, wherein the peptide linker is a poly-Glycine-Serine (G4S) linker, optionally wherein the linker is set forth in any one of SEQ ID NOs. 7-10.

11. The lipid particle of any one of claims 1-10, wherein the universal adapter system comprises a tag component and a catcher component from a SpyTag/SpyCatcher system; a SnoopTag/SnoopCatcher system; a TEFCA tag/catcher system, or a DogTag/DogCatcher system.

12. The lipid particle of any one of claim 1-11, wherein the universal adapter system comprises a SpyTag tag component and a SpyCatcher catcher component selected from the group comprising:

(i) SpyTagΔ and SpyCatcherΔ (e.g., SEQ ID NOs 11 and 16); or

(ii) SpyTag003 and SpyCatcher003 (e.g., SEQ ID NOs 12 and 17).

13. The lipid particle of any one of claim 1-11, wherein the universal adapter system comprises:

(i) a SnoopTag tag component (e.g., SEQ ID NO. 13) and a SnoopCatcher catcher component (e.g., SEQ ID NO. 18);

(ii) a TEFCA tag component (e.g., SEQ ID NO. 14) and a TEFCACatcher catcher component (e.g., SEQ ID NO. 19); or

(iii) a DogTag tag component (e.g., SEQ ID NO. 15) and a DogCatcher catcher component (e.g., SEQ ID NO. 20).

14. The lipid particle of any of claims 1-2, and 4-13, wherein the targeting moiety is specific for a target molecule expressed on the surface of a target cell.

15. The lipid particle of claim 14, wherein the targeting moiety is selected from the group consisting of an antibody or antigen-binding fragment, an engineered binding domain, a nanobody, a DARPin, an Aptamer, an Affimer, an Affibody, a Knottin, an Avimer, a Monobody, an Anticalin, a Fynomer, and a targeting peptide.

16. The lipid particle of any of claims 14 or 15, wherein the targeting moiety is selected from the group consisting of a single domain antibody or a single chain variable fragment (scFv).

17. The lipid particle of claim 16, wherein the single domain antibody is a VHH.

18. The lipid particle of any of claims 1-17, wherein the paramyxovirus envelope attachment protein is an envelope attachment protein from a Nipah virus, Hendra virus, or Measles virus, or is a variant or biologically active portion thereof of any of the foregoing.

19. The lipid particle of claim 18, wherein the paramyxovirus envelope attachment protein is a variant Nipah G protein (NiV-G) that is a variant or a biologically active portion of a wild-type NiV-G.

20. The lipid particle of claim 19, wherein the variant NiV-G tis truncated by up to 40 contiguous amino acids at or near the N-terminus of the wild-type NiV-G set forth in SEQ ID NO:5.

21. The lipid particle of claim 19 or 20, wherein the variant NiV-G has a truncation of amino acids 2-34 of the wild-type NiV-G set forth in SEQ ID NO:5.

22. The lipid particle of any of claims 1-21, wherein the paramyxovirus envelope attachment protein comprises one or more mutations that reduces native tropism relative to the wild-type paramyxovirus envelope attachment protein not comprising the one or more mutations.

23. The lipid particle of any of claims 19-22, wherein the variant NiV-G exhibits reduced binding to Ephrin B2 or Ephrin B3.

24. The lipid particle of claim 13, wherein the variant NiV-G comprises: one or more amino acid substitutions corresponding to amino acid substitutions selected from the group consisting of E501A, W504A, Q530A and E533A with reference to numbering set forth in SEQ ID NO:5.

25. The lipid particle of claim 23 or claim 24, wherein the variant NiV-G comprises amino acid substitutions E501A, W504A, Q530A and E533A with reference to numbering set forth in SEQ ID NO:5.

26. The lipid particle of any of claims 19-25, wherein the variant NiV-G has the amino acid sequence set forth in SEQ ID NO: 6 or an amino acid sequence having at or about 80%, at least at or about 81 %, at least at or about 82%, at least at or about 83%, at or about 84%, at least at or about 85%, at least at or about 86%, or at least at or about 87%, at least at or about 88%, or at least at or about 89%, at least at or about 90%, at least at or about 91 %, at least at or about 92%, at least at or about 93%, at least at or about 94%, at least at or about 95%, at or about 96%, at least at or about 97%, at least at or about 98%, or at least at or about 99% sequence identity to SEQ ID NO:6.

27. The lipid particle of any of claims 19-26, wherein the variant NiV-G has the amino acid sequence set forth in SEQ ID NO:6.

28. The lipid particle of any of claims 1-27, wherein the at least one paramyxovirus fusion (F) protein is a wilt-type Nipah F protein (NiV-F) or is a functionally active variant or a biologically active portion thereof.

29. The lipid particle of claim 28, wherein the NiV-F protein or the biologically active portion is a truncated NiV-F that is truncated by at least 22 amino acids or at least 20 amino acids, at or near the C-terminus of wild-type NiV-F set forth in SEQ ID NO:1.

30. The lipid particle of any of claims 28 or 29, wherein the NiV-F protein or the biologically active portion thereof has a 22 amino acid truncation at or near the C-terminus of the wild-type NiV-F protein.

31. The lipid particle of any of claims 28-30, wherein the NiV-F protein comprises a deletion in it cytoplasmic tail and lacks amino acid residues 525-546 of SEQ ID NO:1.

32. The lipid particle of any of claims 28-31, wherein the NiV-F protein or the biologically active portion thereof has the sequence set forth in SEQ ID NO: 3 or 4, or a sequence of amino acids that exhibits at least at or about 80%, 85%, 90% or 95% sequence identity to the sequence set forth in SEQ ID NO: 3 or 4.

33. The lipid particle of any of claims 1-32, wherein the lipid particle is a vector, optionally a viral vector, optionally wherein the vector is a lentiviral vector.

34. The lipid particle of any of claims 1-33, wherein the lipid particle further comprises an exogenous agent for delivery to the target cell.

35. A combination, comprising:

(a) the lipid particle of any of claims 3-13, and 18-34, wherein the paramyxovirus envelope attachment protein is fused to a tag polypeptide component or a catcher polypeptide component of a universal adapter system, and

(b) a targeting moiety fused to the other component of the universal adapter system.

36. The combination of claim 35, wherein the universal adapter system comprises a tag component and a catcher component from a SpyTag/SpyCatcher system; a SnoopTag/SnoopCatcher system; a TEFCA tag/catcher system, or a DogTag/DogCatcher system.

37. The combination of claim 35 or 36, wherein the universal adapter system comprises a SpyTag tag component and a SpyCatcher catcher component selected from the group comprising:

(i) SpyTagΔ and SpyCatcherΔ (e.g., SEQ ID NOs 11 and 16); or

(ii) SpyTag003 and SpyCatcher003 (e.g., SEQ ID NOs 12 and 17).

38. The combination of any one of claims 35-37, wherein the universal adapter system comprises:

(i) a SnoopTag tag component (e.g., SEQ ID NO. 13) and a SnoopCatcher catcher component (e.g., SEQ ID NO. 18);

(ii) a TEFCA tag component (e.g., SEQ ID NO. 14) and a TEFCACatcher catcher component (e.g., SEQ ID NO. 19); or

(iii) a DogTag tag component (e.g., SEQ ID NO. 15) and a DogCatcher catcher component (e.g., SEQ ID NO. 20).

39. The combination of any of claims 35-38, wherein the targeting moiety is specific for a target molecule expressed on the surface of a target cell.

40. The combination of claim 39, wherein the targeting moiety is selected from the group consisting of an antibody or antigen-binding fragment, an engineered binding domain, a nanobody, a

DARPin, an Aptamer, an Affimer, an Affibody, a Knottin, an Avimer, a Monobody, an Anticalin, a Fynomer, and a targeting peptide.

41. The combination of any of claims 39 or 40, wherein the targeting moiety is selected from the group consisting of a single domain antibody or a single chain variable fragment (scFv).

42. The combination of claim 41, wherein the single domain antibody is a VHH.

43. A method of making a lipid particle comprising a paramyxovirus envelope attachment protein attached to a targeting moiety via a universal adapter system, said system comprising a tag polypeptide component and a catcher polypeptide component, wherein the paramyxovirus envelope attachment protein is fused to one of the components, and the targeting moiety is fused to the other of the components; and at least one paramyxovirus fusion (F) protein, the method comprising:

(a) providing a cell that comprises the lipid particle of any one of claims 3-13 and 18-34;

(b) culturing the cell under condition that allow for production of the lipid particle,

(c) separating, enriching, or purifying the lipid particle from the cell, thereby making the lipid particle;

(d) mixing the lipid particle with a targeting moiety fused to the other component of the universal adapter system; and

(e) removing unbound targeting moiety from the lipid particles to collect lipid particles with the paramyxovirus envelope attachment protein attached to the targeting moiety.

44. The method of claim 43, further comprising concentrating the collected lipid particles.

45. A lipid particle produced by the method of claim 43 or 44.

46. A producer cell, comprising the lipid particle of any one of claims 3-13 and 18-34.

47. A kit comprising the combination of any of claim 35-42.

48. An article of manufacture comprising the kit of claim 47.

49. A method of selective transduction, the method comprising contacting a target cell with the lipid particle of any of claims 1, 2 and 4-34.

50. A method of selective transduction, the method comprising:

exposing a composition comprising a target cell with (1) the lipid particle of any of claims 3-13, and 18-34, wherein the lipid particle comprises a paramyxovirus envelope attachment protein fused to a tag polypeptide component or a catcher polypeptide component of a universal adapter system; and (2) a targeting moiety fused to the other component of the universal adapter system.

51. The method of claim 49 or claim 50 that is performed in vitro or ex vivo.

52. The method of any of claims 49 or claim 50 that is performed in vivo.

53. A method of selective transduction of a target cell in a subject, the method comprising administering the lipid particle of any of claims 1, 2 and 4-34 to a subject.

54. A method of selective transduction of a target cell in a subject, the method comprising:

(a) administering the lipid particle of any of claims 3-13, and 18-34 to a subject, wherein the lipid particle comprises a paramyxovirus envelope attachment protein fused to a tag polypeptide component or a catcher polypeptide component of a universal adapter system; and

(b) administering a targeting moiety fused to the other component of the universal adapter system to the subject.

55. The method of any of claims 50-52 and 54, wherein the universal adapter system comprises a tag component and a catcher component from a SpyTag/SpyCatcher system; a SnoopTag/SnoopCatcher system; a TEFCA tag/catcher system, or a DogTag/DogCatcher system.

56. The method of any of claims 50-52, 54 and 55, wherein the universal adapter system comprises a SpyTag tag component and a SpyCatcher catcher component selected from the group comprising:

(i) SpyTagΔ and SpyCatcherΔ (e.g., SEQ ID NOs 11 and 16); or

(ii) SpyTag003 and SpyCatcher003 (e.g., SEQ ID NOs 12 and 17).

57. The method of any of claims 50-52 and 54-56, wherein the universal adapter system comprises:

(i) a SnoopTag tag component (e.g., SEQ ID NO. 13) and a SnoopCatcher catcher component (e.g., SEQ ID NO. 18);

(ii) a TEFCA tag component (e.g., SEQ ID NO. 14) and a TEFCACatcher catcher component (e.g., SEQ ID NO. 19); or

(iii) a DogTag tag component (e.g., SEQ ID NO. 15) and a DogCatcher catcher component (e.g., SEQ ID NO. 20).

58. The method of any of claims 50-52 and 54-57, wherein the targeting moiety is specific for a target molecule expressed on the surface of the target cell.

59. The method of claim 58, wherein the targeting moiety is selected from the group consisting of an antibody or antigen-binding fragment, an engineered binding domain, a nanobody, a DARPin, an Aptamer, an Affimer, an Affibody, a Knottin, an Avimer, a Monobody, an Anticalin, a Fynomer, and a targeting peptide.

60. The method of any of claims 58 or 59, wherein the targeting moiety is selected from the group consisting of a single domain antibody or a single chain variable fragment (scFv).

61. The method of claim 60, wherein the single domain antibody is a VHH.

62. A paramyxovirus envelope attachment protein with a targeting moiety attached via a universal adapter system, said system comprising a tag polypeptide component and a catcher polypeptide component, wherein the paramyxovirus envelope attachment protein is fused to one of the components, and the targeting moiety is fused to the other of the components.

63. A paramyxovirus envelope attachment protein comprising one component of a universal adapter system with a targeting moiety attached via a universal adapter system, said system comprising a tag polypeptide component and a catcher polypeptide component, wherein the paramyxovirus envelope attachment protein is fused to the one component.

64. The paramyxovirus envelope attachment protein of claim 62 or 63, wherein the component fused to the paramyxovirus envelope attachment protein is linked to the C-terminus of the paramyxovirus envelope attachment protein.

65. The paramyxovirus envelope attachment protein of claim 62 or 63, wherein the component fused to the paramyxovirus envelope attachment protein is inserted into a loop region of the paramyxovirus envelope attachment protein.

66. The paramyxovirus envelope attachment protein of claim 65, wherein the loop is selected from the group consisting of:

(i) amino acid residues 194-197;

(ii) amino acid residues 208-213;

(iii) amino acid residues 241-243;

(iv) amino acid residues 242-243;

(v) amino acid residues 388-391;

(vi) amino acid residues 402-405;

(vii) amino acid residues 489-491; and

(viii) amino acid residues 495-497;

optionally wherein the peptide loop is 388-391, 489-491, or 495-497.

67. The paramyxovirus envelope attachment protein of any of claims 62-66, wherein the paramyxovirus envelope attachment protein is from Nipah.

68. The paramyxovirus envelope attachment protein of claim 67, wherein the paramyxovirus envelope attachment protein is a NiV-G or is a biologically active variant or truncated form thereof.

69. The paramyxovirus envelope attachment protein of any of claims 62-68, wherein the targeting moiety is selected from the group consisting of an antibody or antigen-binding fragment, an engineered binding domain, a nanobody, a DARPin, an Aptamer, an Affimer, an Affibody, a Knottin, an Avimer, a Monobody, an Anticalin, a Fynomer, and a targeting peptide.

FIG. 1B

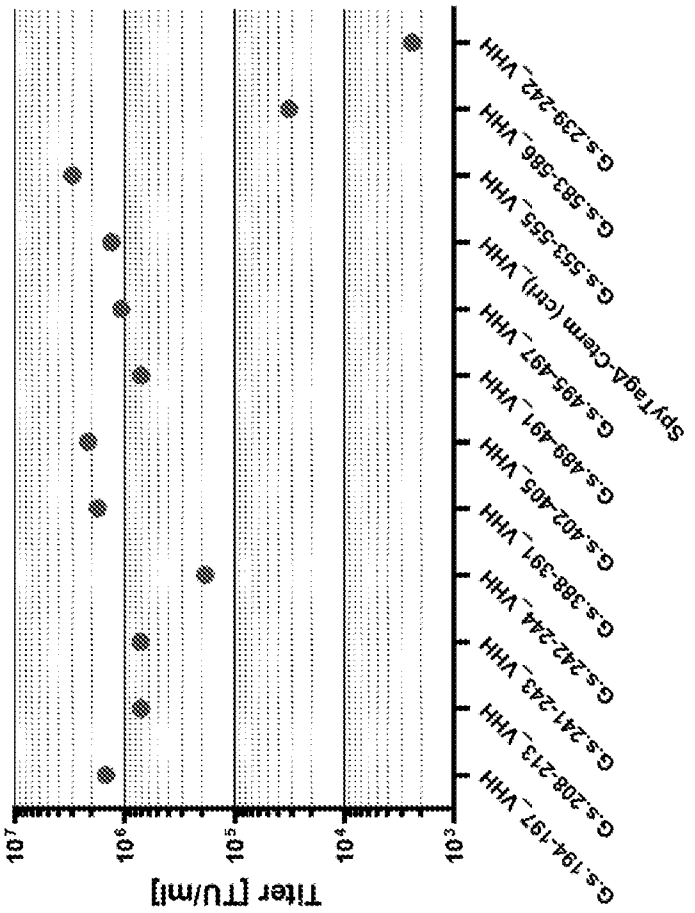

Titer [TU/ml]
$10^7$
$10^6$
$10^5$
$10^4$
$10^3$
G.s.194-197_VHH
G.s.208-213_VHH
G.s.241-243_VHH
G.s.242-244_VHH
G.s.388-391_VHH
G.s.402-405_VHH
G.s.489-491_VHH
G.s.495-497_VHH
SpyTag-Cterm (ctrl)_VHH
G.s.553-555_VHH
G.s.583-586_VHH
G.s.239-242_VHH

FIG. 1A

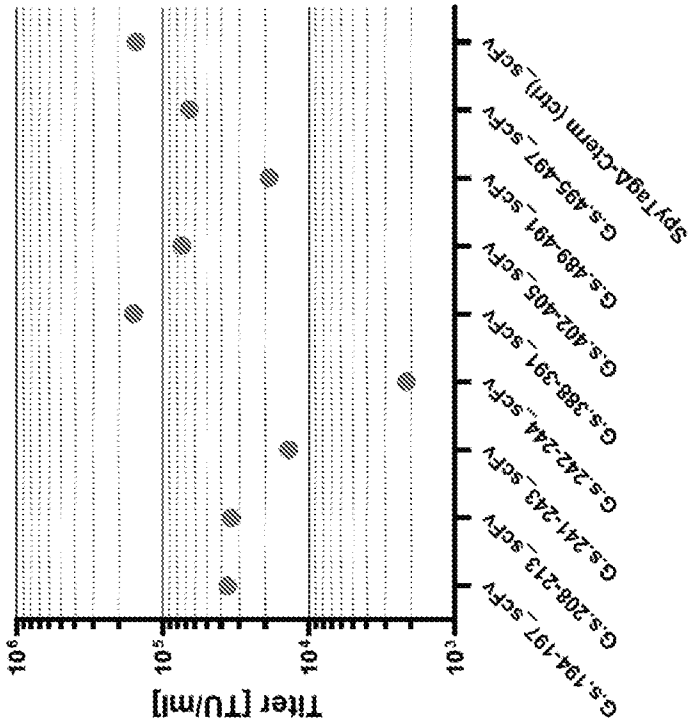

Titer [TU/ml]
$10^6$
$10^5$
$10^4$
$10^3$
G.s.194-197_scFv
G.s.208-213_scFv
G.s.241-243_scFv
G.s.242-244_scFv
G.s.388-391_scFv
G.s.402-405_scFv
G.s.489-491_scFv
G.s.495-497_scFv
SpyTag-Cterm (ctrl)_scFv

FIG. 1D

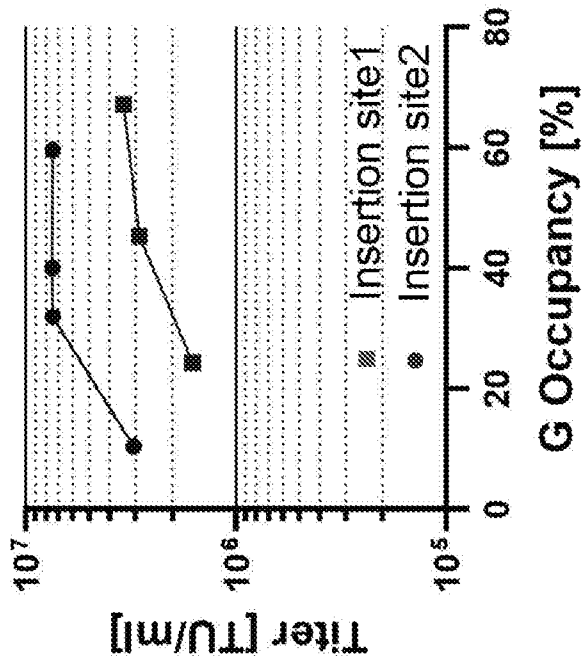

Titer [TU/ml]
$10^7$
$10^6$
$10^5$
0
20
40
60
80
G Occupancy [%]
Insertion site1
Insertion site2

FIG. 1C

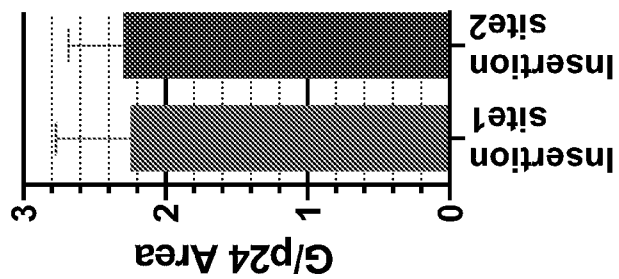
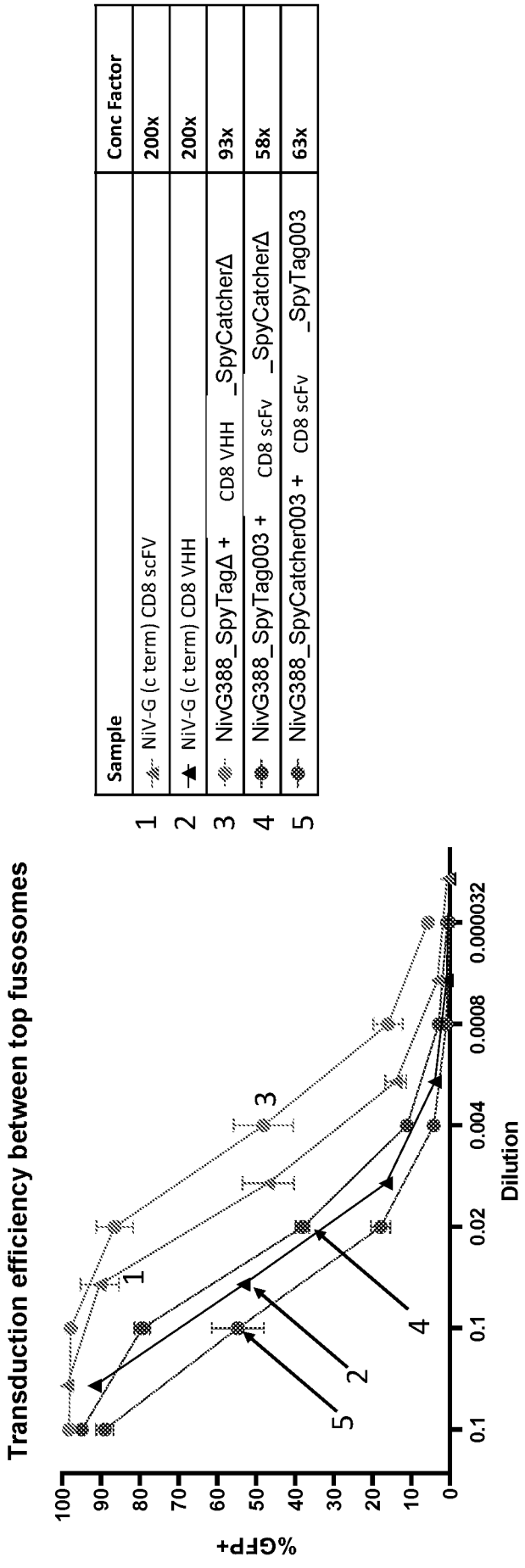

G/p24 Area
0
1
2
3
Insertion site1
Insertion site2

FIG. 2




| | Sample | Conc Factor |
|---|---|---|
| 1 | NiV-G (c term) CD8 scFV | 200x |
| 2 | NiV-G (c term) CD8 VHH | 200x |
| 3 | NivG388_SpyTagΔ + CD8 VHH _SpyCatcherΔ | 93x |
| 4 | NivG388_SpyTag003 + CD8 scFv _SpyCatcherΔ | 58x |
| 5 | NivG388_SpyCatcher003 + CD8 scFv _SpyTag003 | 63x |

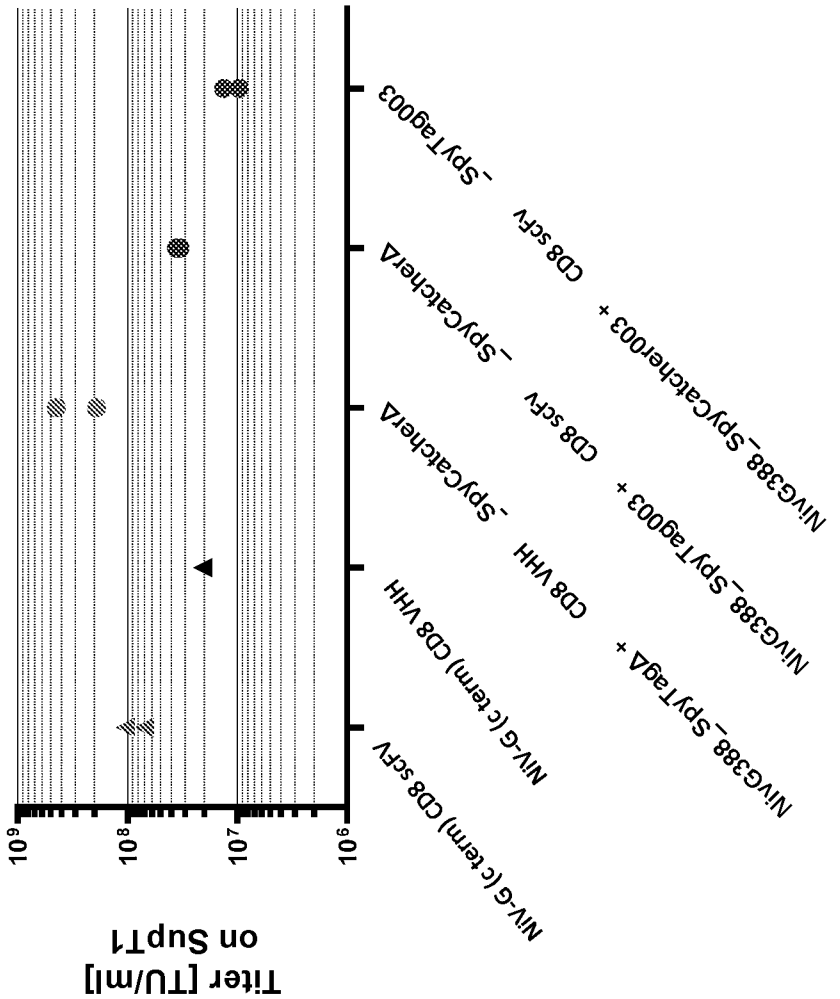

Titer [TU/ml]
on SupT1
$10^9$
$10^8$
$10^7$
$10^6$
NiV-G (c.term) CD8 scFv
NiV-G (c.term) CD8 VHH
NiVG388_SpyTag A + CD8 VHH _SpyCatcherA
NiVG388_SpyTag003 + CD8 scFv _SpyCatcherA
NiVG388_SpyCatcher003 + CD8 scFv _SpyTag003

FIG. 3A

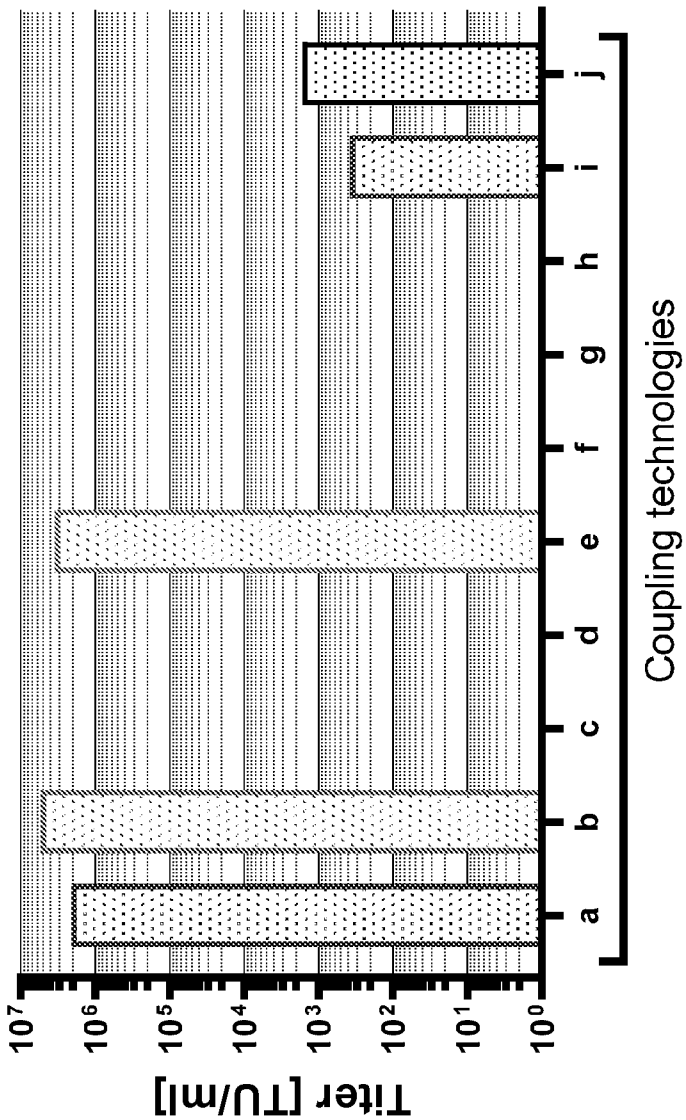

Titer [TU/ml]
10^7
10^6
10^5
10^4
10^3
10^2
10^1
10^0
a
b
c
d
e
f
g
h
i
j
Coupling technologies

FIG. 3B

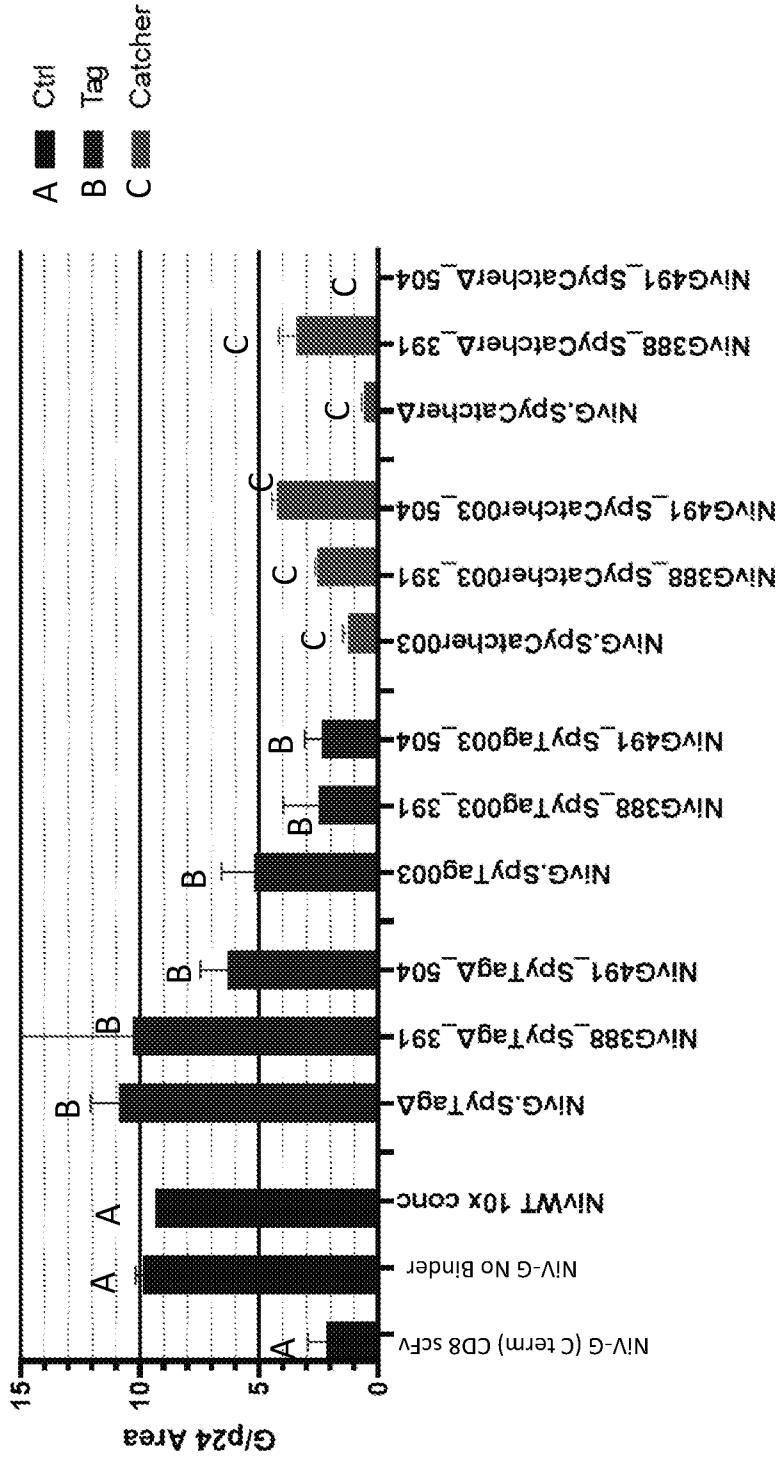

A Ctrl
B Tag
C Catcher
G/p24 Area
0
5
10
15
NiV-G (C term) CD8 scFv
NiV-G No Binder
NiVWT 10x conc
NiVG.SpyTagA
NiVG388_SpyTagA_391
NiVG491_SpyTagA_504
NiVG.SpyTag003
NiVG388_SpyTag003_391
NiVG491_SpyTag003_504
NiVG.SpyCatcher003
NiVG388_SpyCatcher003_391
NiVG491_SpyCatcher003_504
NiVG.SpyCatcherΔ
NiVG388_SpyCatcherΔ_391
NiVG491_SpyCatcherΔ_504

FIG. 4

FIG. 5A

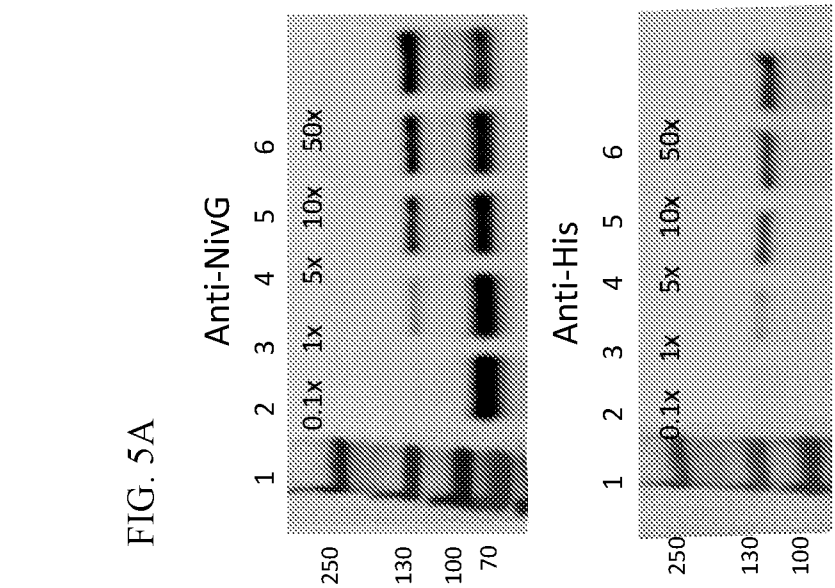

Anti-NivG
1
2
0.1x
3
1x
4
5x
5
10x
6
50x
250
130
100
70
Anti-His
1
2
0.1x
3
1x
4
5x
5
10x
6
50x
250
130
100

FIG. 5B

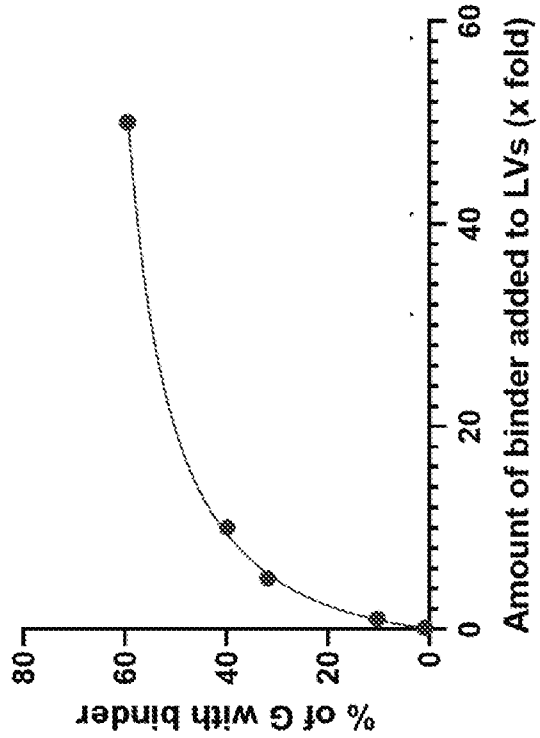

Occupancy
% of G with binder
0
20
40
60
80
0
20
40
60
Amount of binder added to LVs (x fold)

FIG. 5D

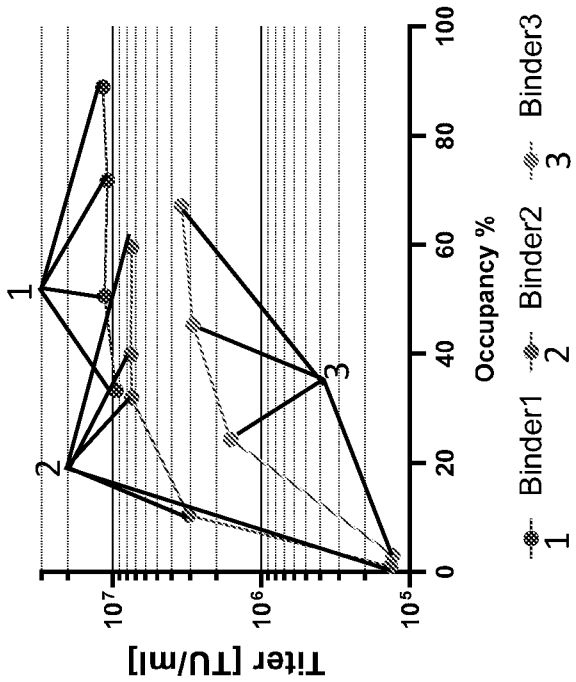

Titer [TU/ml]
$10^7$
$10^6$
$10^5$
0
20
40
60
80
100
Occupancy %
1
2
3
Binder1
Binder2
Binder3

FIG. 5C

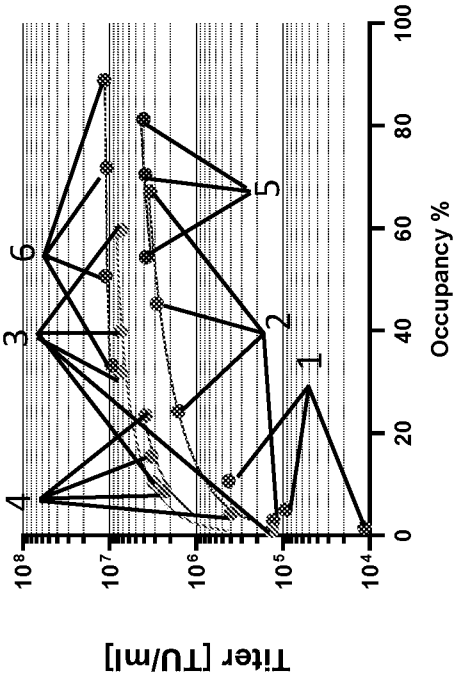

Titer [TU/ml]
$10^8$
$10^7$
$10^6$
$10^5$
$10^4$
0
20
40
60
80
100
Occupancy %
1
2
3
4
5
6
VHH
VHH-1k
scFv
scFv-2
scFv-2c
scFv-1k

FIG. 6B

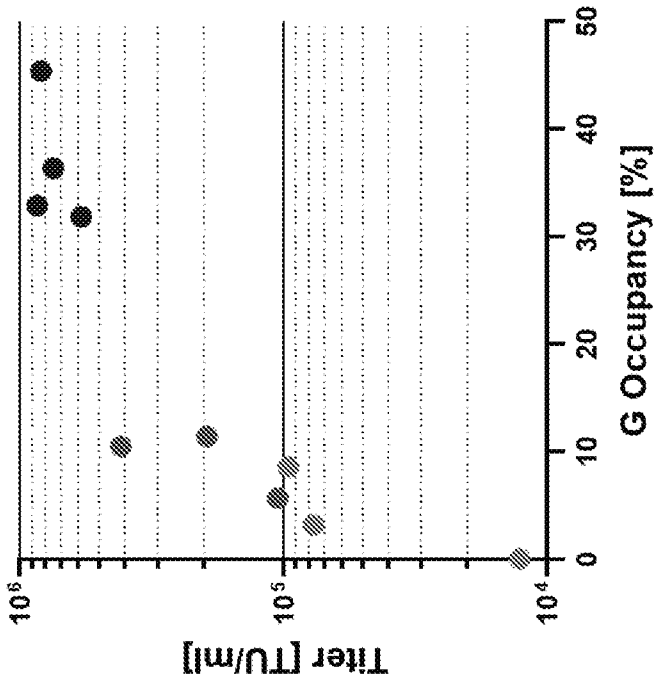
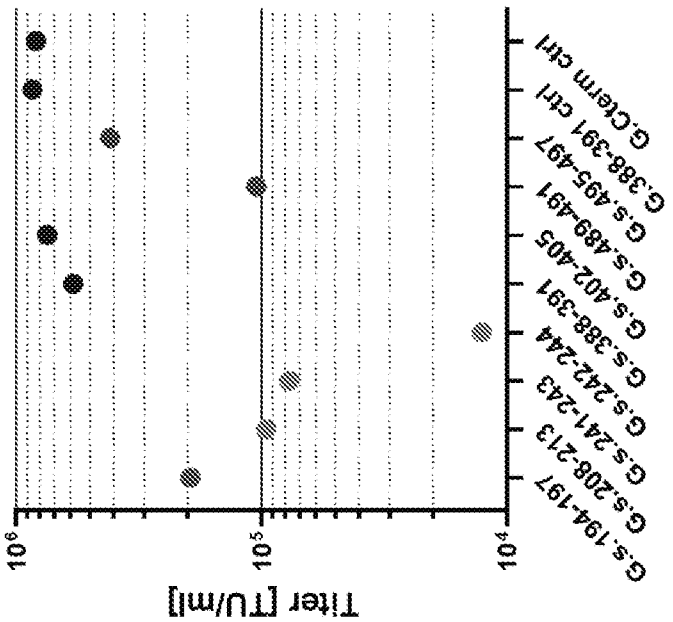

Titer [TU/ml]
$10^6$
$10^5$
$10^4$
0
10
20
30
40
50
G Occupancy [%]

FIG. 6A


Titer [TU/ml]
$10^6$
$10^5$
$10^4$
G.s.194-197
G.s.208-213
G.s.241-243
G.s.242-244
G.s.388-391
G.s.402-405
G.s.489-491
G.s.495-497
G.388-391 ctrl
G.Cterm ctrl

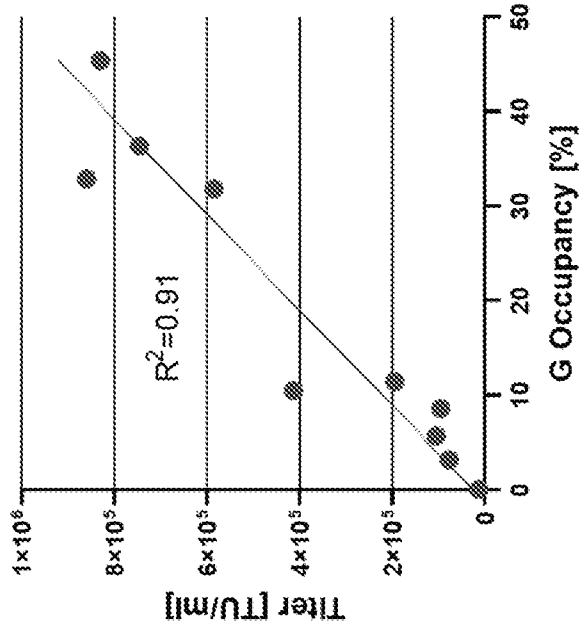

Titer [TU/ml]
1×10^6
8×10^5
6×10^5
4×10^5
2×10^5
0
R^2=0.91
0
10
20
30
40
50
G Occupancy [%]

FIG. 6C

FIG. 7

MKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQNYTRSTDNQAVIKDALQGIQQQIKGLADKIGTEIGP
KVSLIDTSSTITIPANIGLLGSKISQSTASINENVNEKCKFTLPPLKIHECNISCPNPLPFREYRPQTEGVS**NLVG**
**LPNNICLQKTSNQILKPKLI**SYTL**PVVGQSGT**CITDPLLA**MDEG**YFAYSHLER**IGSCSRGV**SKQRIIGVGEVL
**DRGDEVP**SLFMTNV**WTPPNPNTV**YHCSAVY**NNEF**YYVLCA**VSTVGDPILNSTYWSGSL**MMTRL**AVKP**
**KSNG**GGYNQHQ**LALRSIEKGRYDK**VMPYGPSGIK**QGDT**LYFPAVGFL**VRTEFKYNDSNCPITKCQYSKP**E
NCRLS**MGIRPNSHY**ILRSGLLKY**NLSDGENPKV**VFI**EISDQRLSI**GSPSKIYD**SLGQ**PVFYQASFSWDTMIK
FGDVL**TVNPLV**VNWRNNTVISRP**GQSQCPRFNTCPAICAEG**VYNDAFLI**DRINWI**SAGVF**LDSNATAAN**
PVFTVF**KDNEI**LYRA**QLASEDTNA**QKTITNCFL**LKNK**IWCISLVEIY**DTGDNV**IRPKLFAV**KIPEQCT**

(SEQ ID NO: 6)

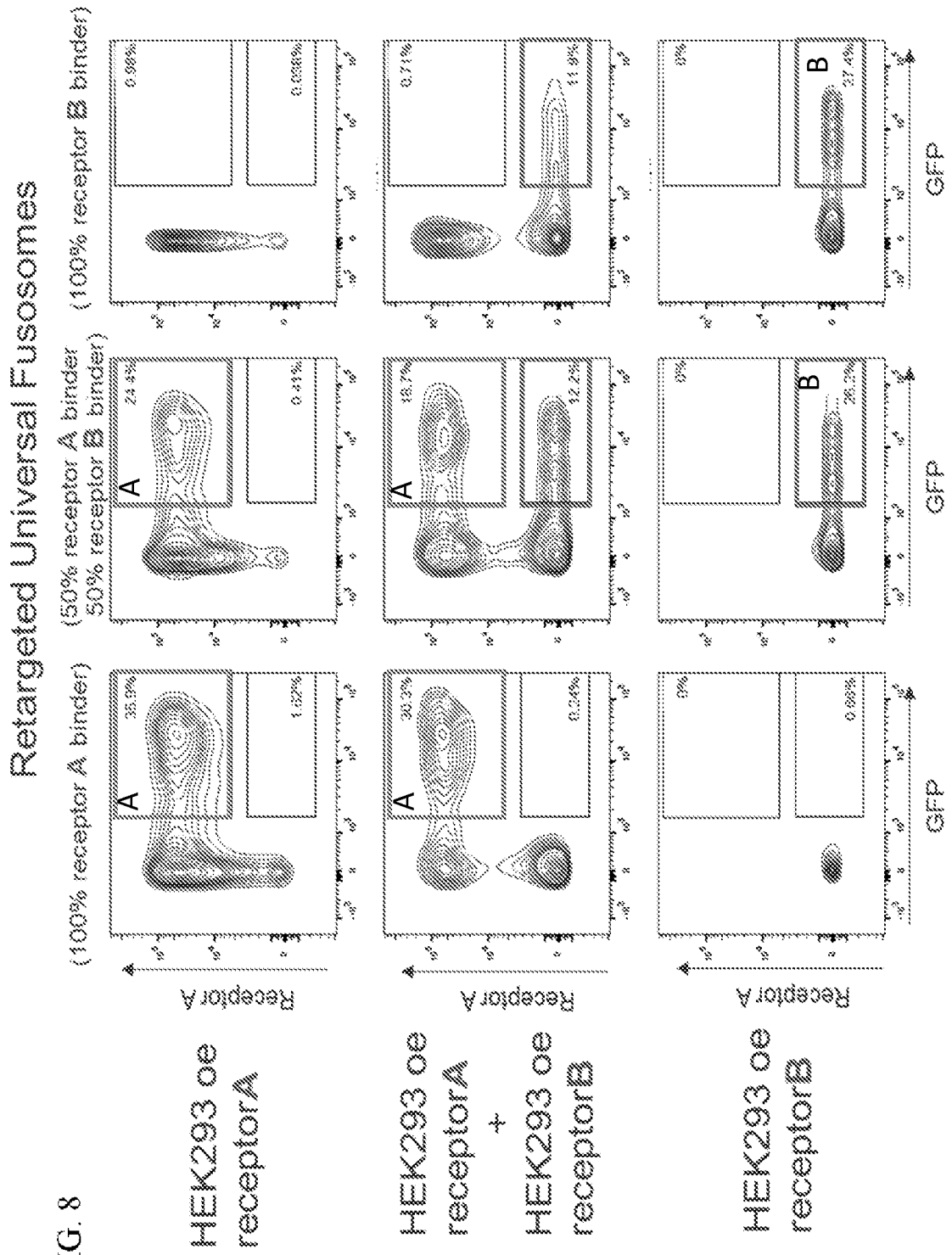

Retargeted Universal Fusosomes
(100% receptor A binder)
(50% receptor A binder
50% receptor B binder)
(100% receptor B binder)
HEK293 oe receptorA
HEK293 oe receptorA + HEK293 oe receptorB
HEK293 oe receptorB
Receptor A
GFP
A
B
35.8%
1.02%
24.4%
0.41%
0%
0.66%
30.3%
0.24%
18.7%
12.2%
0%
26.2%
0.98%
0.038%
0.71%
11.9%
0%
27.4%

FIG. 8

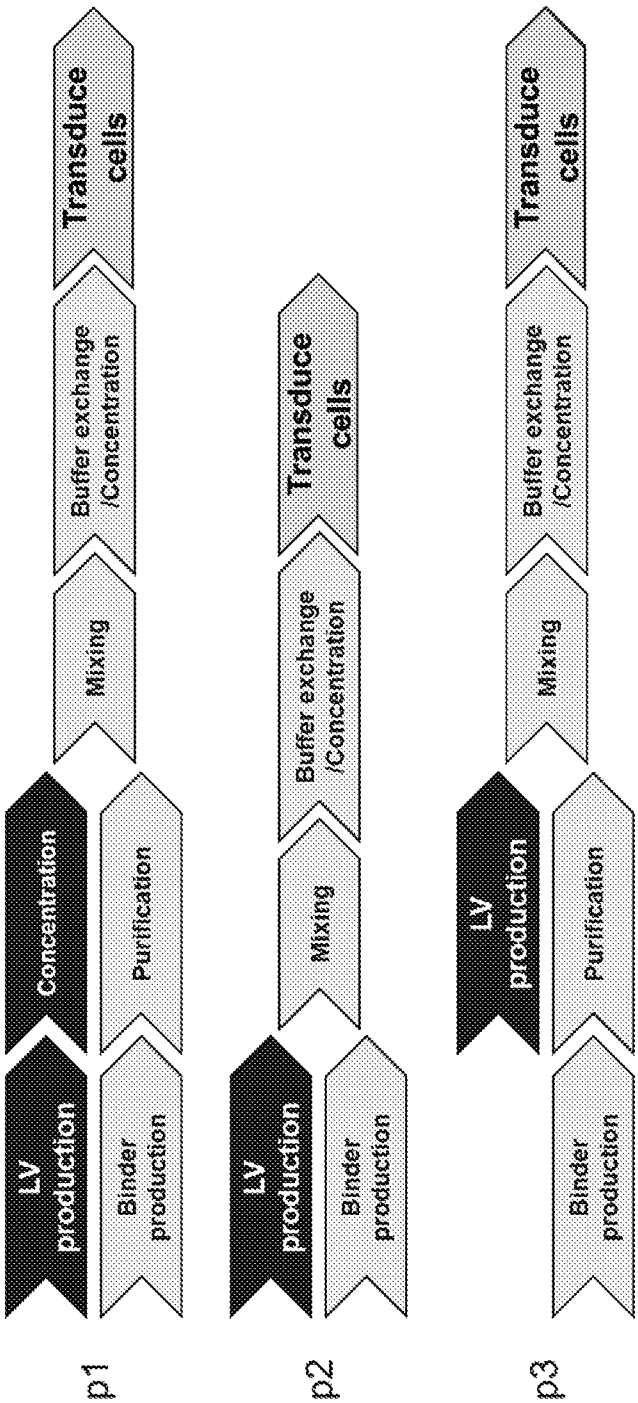
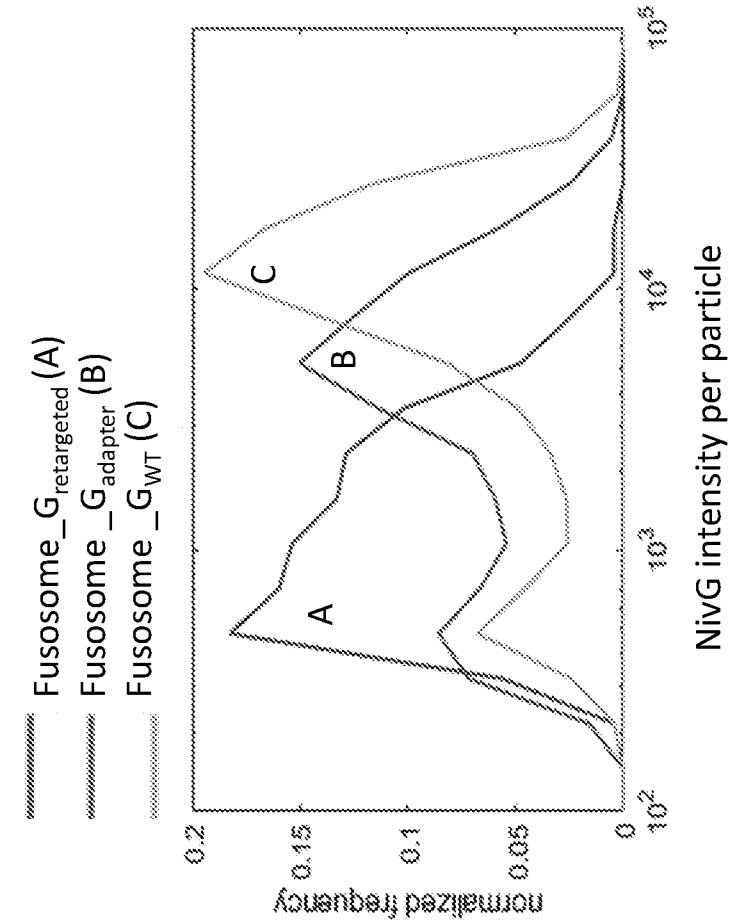

LV production
Concentration
Binder production
Purification
Mixing
Buffer exchange /Concentration
Transduce cells
p1
LV production
Binder production
Mixing
Buffer exchange /Concentration
Transduce cells
p2
LV production
Binder production
Purification
Mixing
Buffer exchange /Concentration
Transduce cells
p3

FIG. 9


Fusosome_G retargeted (A)
Fusosome _G adapter (B)
Fusosome _G WT (C)
normalized frequency
NivG intensity per particle
A
B
C

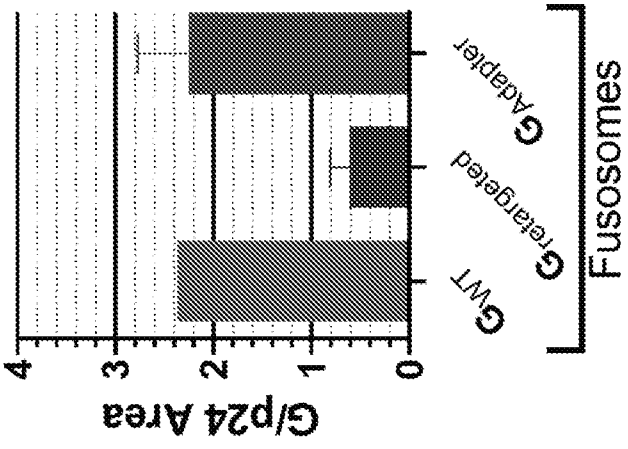

G/p24 Area
G WT
G retargeted
G Adapter
Fusosomes

FIG. 10

FIG. 11B

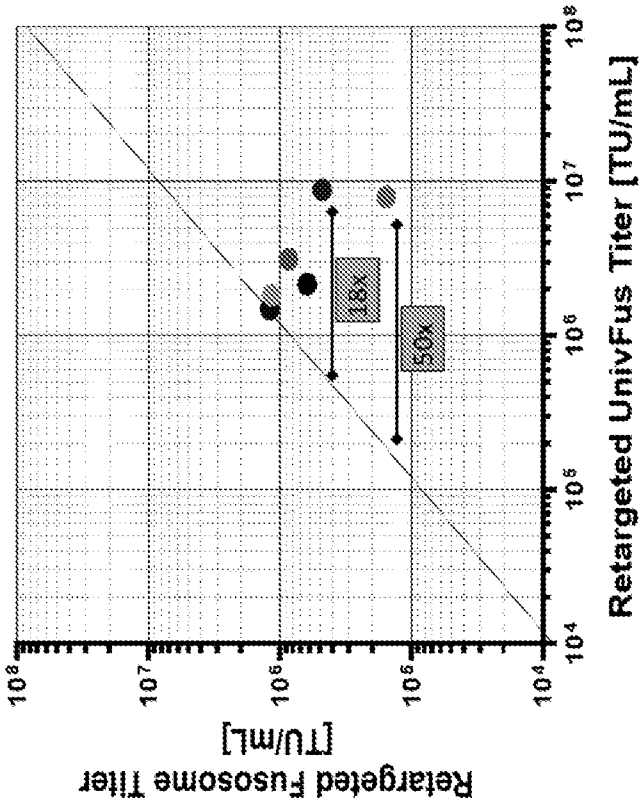
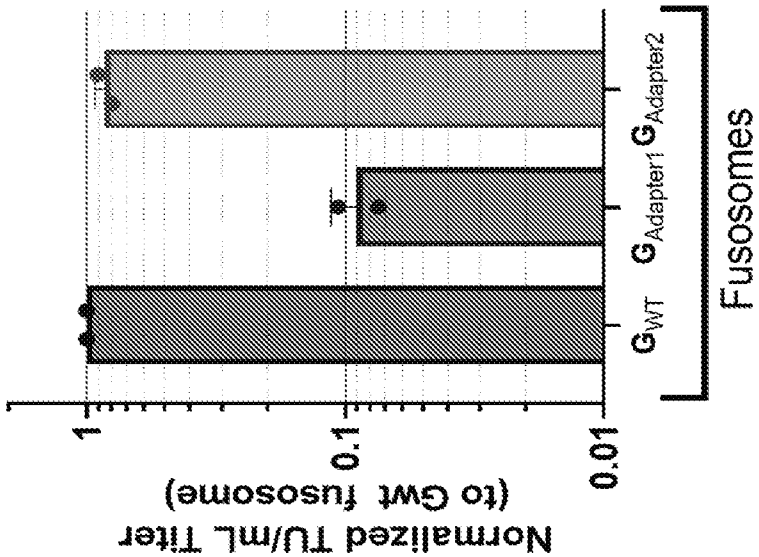



FIG. 11A



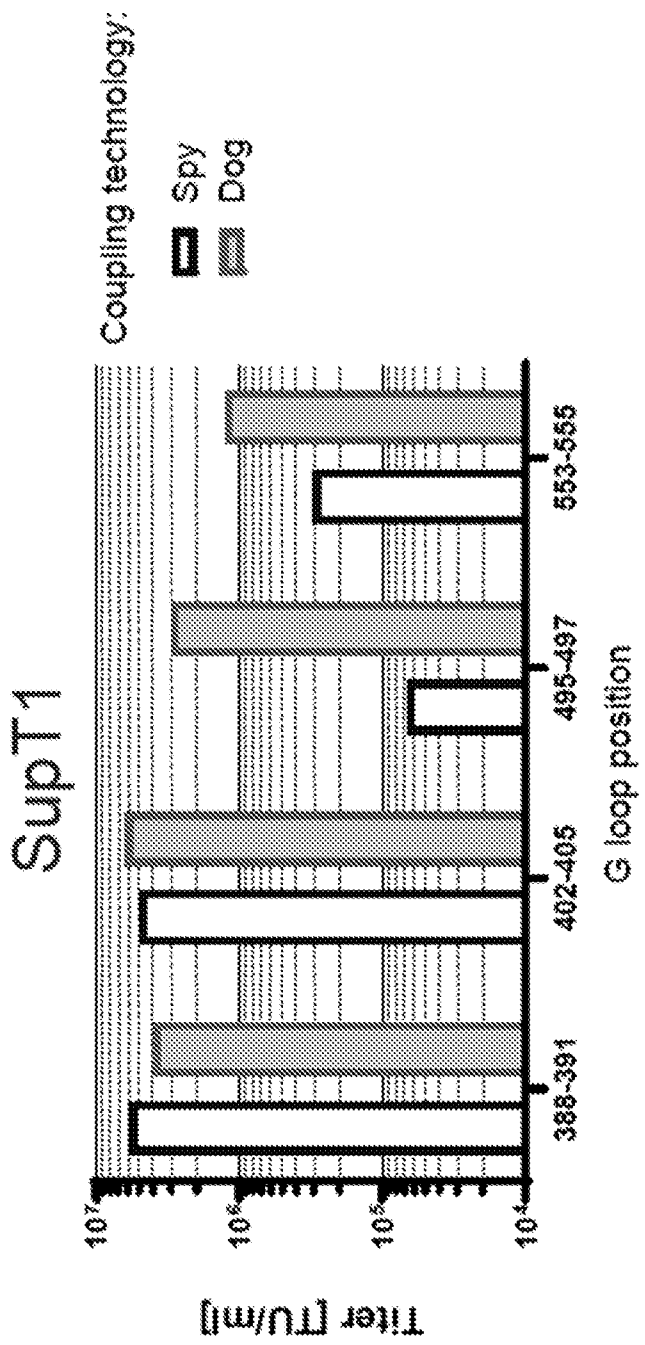

SupT1
Coupling technology:
Spy
Dog
Titer [TU/ml]
$10^7$
$10^6$
$10^5$
$10^4$
388-391
402-405
495-497
553-555
G loop position

FIG. 12

# INTERNATIONAL SEARCH REPORT

International application No
PCT/US2024/025039

**A. CLASSIFICATION OF SUBJECT MATTER**
INV. C12N15/86
ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)
C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
EPO-Internal

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

| Category* | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| Y | KASARANENI NAGARJUN ET AL: "Retargeting Lentiviruses via SpyCatcherSpyTag Chemistry for Gene Delivery into Specific Cell Types", MBIO, vol. 8, no. 6, 12 December 2017 (2017-12-12), pages 1-12, XP093138146, US ISSN: 2161-2129, DOI: 10.1128/mBio abstract figure 1 | 1-69 |
| | ----- | |
| Y | WO 2021/202604 A1 (SANA BIOTECHNOLOGY INC [US] ET AL.) 7 October 2021 (2021-10-07) example 2 | 1-69 |
| | ----- | |
| | -/-- | |

[X] Further documents are listed in the continuation of Box C.

[X] See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance;; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance;; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

| Date of the actual completion of the international search | Date of mailing of the international search report |
|---|---|
| 30 August 2024 | 16/09/2024 |

| Name and mailing address of the ISA/ | Authorized officer |
|---|---|
| European Patent Office, P.B. 5818 Patentlaan 2<br>NL - 2280 HV Rijswijk<br>Tel. (+31-70) 340-2040,<br>Fax: (+31-70) 340-3016 | Mauhin, Viviane |

# INTERNATIONAL SEARCH REPORT

International application No
PCT/US2024/025039

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

| Category* | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| Y | WO 2017/182585 A1 (ECOLE NORMALE SUPERIEURE LYON [FR]; CENTRE NAT RECH SCIENT [FR] ET AL.)<br>26 October 2017 (2017-10-26)<br>example 1 | 1-69 |
| Y | WO 2022/216915 A1 (SANA BIOTECHNOLOGY INC [US]; CRUITE PATRICIA [US] ET AL.)<br>13 October 2022 (2022-10-13)<br>example 1 | 1-69 |
| A | KEEBLE ANTHONY H ET AL: "DogCatcher allows loop-friendly protein-protein ligation",<br>CELL CHEMICAL BIOLOGY, ELSEVIER, AMSTERDAM, NL,<br>vol. 29, no. 2, 28 July 2021 (2021-07-28), page 339, XP086965342,<br>ISSN: 2451-9456, DOI: 10.1016/J.CHEMBIOL.2021.07.005<br>[retrieved on 2021-07-28]<br>the whole document | 1-69 |
| A | ANDERSSON ANNE-MARIE C. ET AL: "SnoopLigase peptide-peptide conjugation enables modular vaccine assembly",<br>SCIENTIFIC REPORTS,<br>vol. 9, no. 1,<br>1 December 2019 (2019-12-01), page 4625, XP055879626,<br>DOI: 10.1038/s41598-019-40985-w<br>Retrieved from the Internet: URL:https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6420506/pdf/41598_2019_Article_40985.pdf><br>the whole document | 1-69 |
| A | BULDUN CAN M. ET AL: "SnoopLigase Catalyzes Peptide-Peptide Locking and Enables Solid-Phase Conjugate Isolation",<br>JOURNAL OF THE AMERICAN CHEMICAL SOCIETY,<br>vol. 140, no. 8,<br>28 February 2018 (2018-02-28), pages 3008-3018, XP055940053,<br>ISSN: 0002-7863, DOI: 10.1021/jacs.7b13237<br>the whole document | 1-69 |

-/--

# INTERNATIONAL SEARCH REPORT

International application No
PCT/US2024/025039

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

| Category* | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| A | KEEBLE ANTHONY H. ET AL: "Power to the protein: enhancing and combining activities using the Spy toolbox", CHEMICAL SCIENCE, vol. 11, no. 28, 1 January 2020 (2020-01-01), pages 7281-7291, XP055793574, United Kingdom ISSN: 2041-6520, DOI: 10.1039/D0SC01878C Retrieved from the Internet: URL:https://pubs.rsc.org/en/content/articlepdf/2020/sc/d0sc01878c> the whole document ----- | 1-69 |

# INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2024/025039

**Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item 1.c of the first sheet)**

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing:

   a. [X] forming part of the international application as filed.

   b. [ ] furnished subsequent to the international filing date for the purposes of international search (Rule 13*ter*.1(a)).

      [ ] accompanied by a statement to the effect that the sequence listing does not go beyond the disclosure in the international application as filed.

2. [ ] With regard to any nucleotide and/or amino acid sequence disclosed in the international application, this report has been established to the extent that a meaningful search could be carried out without a WIPO Standard ST.26 compliant sequence listing.

3. Additional comments:

# INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/US2024/025039

| Patent document cited in search report | Publication date | Patent family member(s) | Publication date |
|---|---|---|---|
| WO 2021202604 A1 | 07-10-2021 | AU 2021248815 A1 | 13-10-2022 |
| | | CA 3178308 A1 | 07-10-2021 |
| | | CN 116096866 A | 09-05-2023 |
| | | EP 4127144 A1 | 08-02-2023 |
| | | IL 296621 A | 01-11-2022 |
| | | JP 2023521663 A | 25-05-2023 |
| | | KR 20230006819 A | 11-01-2023 |
| | | US 2021353543 A1 | 18-11-2021 |
| | | WO 2021202604 A1 | 07-10-2021 |
| WO 2017182585 A1 | 26-10-2017 | AU 2017253364 A1 | 15-11-2018 |
| | | AU 2020203813 A1 | 02-07-2020 |
| | | AU 2023258359 A1 | 18-01-2024 |
| | | BR 112018071584 A2 | 12-03-2019 |
| | | CA 3021549 A1 | 26-10-2017 |
| | | CN 109642243 A | 16-04-2019 |
| | | CN 117126817 A | 28-11-2023 |
| | | EA 201892369 A1 | 31-05-2019 |
| | | EP 3235908 A1 | 25-10-2017 |
| | | EP 3445862 A1 | 27-02-2019 |
| | | EP 4151737 A1 | 22-03-2023 |
| | | ES 2961976 T3 | 14-03-2024 |
| | | IL 262458 A | 31-12-2018 |
| | | JP 7412885 B2 | 15-01-2024 |
| | | JP 2019514373 A | 06-06-2019 |
| | | JP 2021184767 A | 09-12-2021 |
| | | JP 2022087199 A | 09-06-2022 |
| | | JP 2024028379 A | 04-03-2024 |
| | | KR 20190046713 A | 07-05-2019 |
| | | SG 11201809164T A | 29-11-2018 |
| | | US 2019144885 A1 | 16-05-2019 |
| | | US 2023348934 A1 | 02-11-2023 |
| | | WO 2017182585 A1 | 26-10-2017 |
| | | ZA 201807161 B | 26-06-2019 |
| WO 2022216915 A1 | 13-10-2022 | AU 2022253020 A1 | 26-10-2023 |
| | | CA 3216353 A1 | 13-10-2022 |
| | | EP 4320242 A1 | 14-02-2024 |
| | | IL 307544 A | 01-12-2023 |
| | | KR 20240028975 A | 05-03-2024 |
| | | TW 202305138 A | 01-02-2023 |
| | | US 2022333134 A1 | 20-10-2022 |
| | | US 2023257773 A1 | 17-08-2023 |
| | | WO 2022216915 A1 | 13-10-2022 |