



(51) International Patent Classification:

A61K 35/00 (2006.01) A61K 39/00 (2006.01)

A61K 35/76 (2015.01) A61K 39/12 (2006.01)

A61K 38/04 (2006.01) A61K 48/00 (2006.01)

(21) International Application Number:

PCT/US2021/025828

(22) International Filing Date:

05 April 2021 (05.04.2021)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

63/005,160 03 April 2020 (03.04.2020) US

(71) Applicant: GRITSTONE BIO, INC. [US/US]; 5959 Horton Street, Suite 300, Emeryville, California 94608 (US).

(72) Inventors: FERGUSON, Andrew; c/o Gritstone Oncology, INC., 5959 Horton Street, Suite 300, Emeryville, California 94608 (US). ROUSSEAU, Raphael; c/o Gritstone Oncology, INC., 5959 Horton Street, Suite 300, Emeryville, California 94608 (US). YELENSKY, Roman; c/o Gritstone Oncology, INC., 5959 Horton Street, Suite 300,

Emeryville, California 94608 (US). SUN, James Xin; c/o Gritstone Oncology, INC., 5959 Horton Street, Suite 300, Emeryville, California 94608 (US). DAVIS, Matthew Joseph; c/o Gritstone Oncology, INC., 5959 Horton Street, Suite 300, Emeryville, California 94608 (US). JOOSS, Karin; c/o Gritstone Oncology, INC., 5959 Horton Street, Suite 300, Emeryville, California 94608 (US). RAPPAPORT, Amy Rachel; c/o Gritstone Oncology, INC., 5959 Horton Street, Suite 300, Emeryville, California 94608 (US). SCALLAN, Ciaran Daniel; c/o Gritstone Oncology, INC., 5959 Horton Street, Suite 300, Emeryville, California 94608 (US). GITLIN, Leonid; c/o Gritstone Oncology, INC., 5959 Horton Street, Suite 300, Emeryville, California 94608 (US). PALMER, Christine Denise; c/o Gritstone Oncology, INC., 5959 Horton Street, Suite 300, Emeryville, California 94608 (US).

(74) Agent: NEWMAN, Zachary R. et al.; Goodwin Procter LLP, 100 Northern Avenue, IP Docketing Dept./7TH FL, Boston, Massachusetts 02210 (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, CZ, DE, DJ, DK, DM, DO,

(54) Title: INFECTIOUS DISEASE ANTIGENS AND VACCINES

VEE-MAG25mer srRNA-LNP1(30 µg)

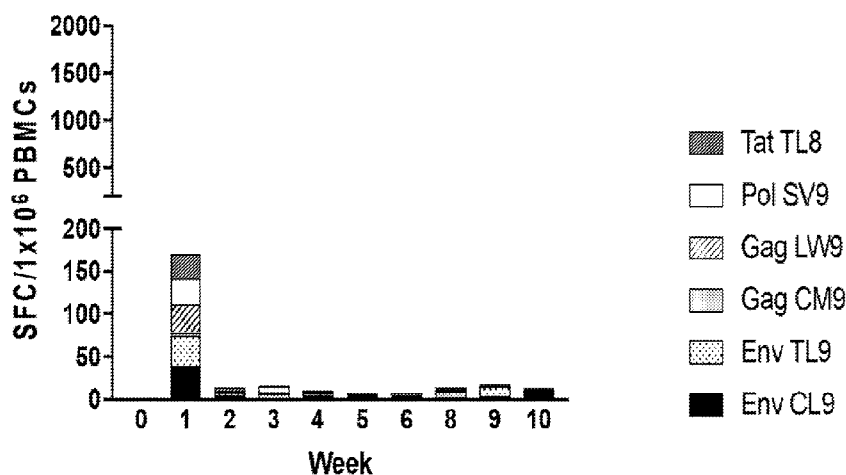


FIG 1A

(57) **Abstract:** Disclosed herein are compositions that include antigen-encoding nucleic acid sequences and/or antigen peptides. Also disclosed are nucleotides, cells, and methods associated with the compositions including their use as vaccines, including vectors and methods for a heterologous prime/boost vaccination strategy.

DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, IT, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, 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, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, 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, 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).

Declarations under Rule 4.17:

- *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*
- *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))*
- *of inventorship (Rule 4.17(iv))*

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))*

INFECTIOUS DISEASE ANTIGENS AND VACCINES

CROSS REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit of U.S. Provisional Application No. 63/005,160 filed April 3, 2020, which is hereby incorporated in its entirety by reference for all purposes.

SEQUENCE LISTING

[0002] The instant application contains a Sequence Listing which has been submitted electronically in ASCII format and is hereby incorporated by reference in its entirety. Said ASCII copy, created on April 5, 2021 is named GSO_090WO_sequencelisting.txt and is 422,179 bytes in size.

BACKGROUND

[0003] One question for antigen vaccine design is which of the many coding mutations present generate the “best” therapeutic antigens, e.g., antigens that can elicit immunity.

[0004] In addition to the challenges of current antigen prediction methods, certain challenges also exist with the available vector systems that can be used for antigen delivery in humans, many of which are derived from humans. For example, many humans have pre-existing immunity to human viruses as a result of previous natural exposure, and this immunity can be a major obstacle to the use of recombinant human viruses for antigen delivery for vaccination, such as infectious disease vaccines.

SUMMARY

[0005] Disclosed herein is a composition for delivery of a self-replicating alphavirus-based expression system, wherein the composition for delivery of the self-replicating alphavirus-based expression system comprises: (A) the self-replicating alphavirus-based expression system, wherein the self-replicating alphavirus-based expression system comprises one or more vectors, wherein the one or more vectors comprises: (a) an RNA alphavirus backbone, wherein the RNA alphavirus backbone comprises: (i) at least one promoter nucleotide sequence, and (ii) at least one polyadenylation (poly(A)) sequence; and (b) a cassette, wherein the cassette comprises: (i) at least one antigen-encoding nucleic acid sequence comprising: a. a nucleic acid sequence encoding an infectious disease organism peptide selected from the group consisting of: a pathogen-derived peptide, a virus-derived peptide, a bacteria-derived peptide, a

fungus-derived peptide, and a parasite-derived peptide, b. optionally a 5' linker sequence, and c. optionally a 3' linker sequence; (ii) optionally, a second promoter nucleotide sequence operably linked to the at least one antigen-encoding nucleic acid sequence; and (iii) optionally, at least one second poly(A) sequence, wherein the second poly(A) sequence is a native poly(A) sequence or an exogenous poly(A) sequence to the alphavirus, and (B) a lipid-nanoparticle (LNP), wherein the LNP encapsulates the self-replicating alphavirus-based expression system.

[0006] Also disclosed herein is a composition for delivery of a self-replicating alphavirus-based expression system, wherein the composition for delivery of the self-replicating alphavirus-based expression system comprises: (A) the self-replicating alphavirus-based expression system, wherein the self-replicating alphavirus-based expression system comprises one or more vectors, wherein the one or more vectors comprises: (a) an RNA alphavirus backbone, wherein the RNA alphavirus backbone comprises the nucleic acid sequence set forth in SEQ ID NO:6, wherein the RNA alphavirus backbone sequence comprises a 26S promoter nucleotide sequence and a poly(A) sequence, wherein the 26S promoter sequence is endogenous to the RNA alphavirus backbone, and wherein the poly(A) sequence is endogenous to the RNA alphavirus backbone; and (b) a cassette integrated between the 26S promoter nucleotide sequence and the poly(A) sequence, wherein the cassette is operably linked to the 26S promoter nucleotide sequence, and wherein the cassette comprises at least one antigen-encoding nucleic acid sequence comprising: a. a nucleic acid sequence encoding an infectious disease organism peptide selected from the group consisting of: a pathogen-derived peptide, a virus-derived peptide, a bacteria-derived peptide, a fungus-derived peptide, and a parasite-derived peptide b. optionally a 5' linker sequence, and c. optionally a 3' linker sequence; and (B) a lipid-nanoparticle (LNP), wherein the LNP encapsulates the self-replicating alphavirus-based expression system.

[0007] In some aspects, the nucleic acid sequence encoding the peptide comprises an epitope-encoding nucleic acid sequence. In some aspects, the nucleic acid sequence encoding the peptide comprises two or more distinct epitope-encoding nucleic acid sequences. In some aspects, the nucleic acid sequence encoding the peptide comprises two or more distinct epitope-encoding nucleic acid sequences. In some aspects, the nucleic acid sequence encoding the peptide comprises between 1-10, between 1-20, between 1-30, between 1-40, between 1-50, between 1-100, between 1-200, between 1-300, between 1-400, or between 1-500 distinct epitope-encoding nucleic acid sequences distinct epitope-encoding nucleic acid sequences. In some aspects, the nucleic acid sequence encoding the peptide comprises between 2-10, between

2-20, between 2-30, between 2-40, between 2-50, between 2-100, between 2-200, between 2-300, between 2-400, or between 2-500 distinct epitope-encoding nucleic acid sequences distinct epitope-encoding nucleic acid sequences. In some aspects, the nucleic acid sequence encoding the peptide comprises a nucleic acid sequence encoding a full-length protein expressed in the infectious disease organism. In some aspects, the nucleic acid sequence encoding the peptide comprises a nucleic acid sequence encoding a protein domain of the protein expressed in the infectious disease organism. In some aspects, the nucleic acid sequence encoding the peptide comprises a nucleic acid sequence encoding a protein subunit of the protein expressed in the infectious disease organism. In some aspects, the nucleic acid sequence encoding the peptide comprises two or more distinct nucleic acid sequences encoding a peptide selected from: an epitope, a full-length protein, a protein subunit, a protein domain, and combinations thereof of the protein expressed in the infectious disease organism.

[0008] In some aspects, the encoded peptide or peptides is capable of stimulating an immune response when expressed in a subject. In some aspects, the encoded peptide or peptides is capable of stimulating a T cell response when expressed in a subject. In some aspects, the encoded peptide or peptides is capable of stimulating a B cell response when expressed in a subject. In some aspects, the encoded peptide or peptides is capable of stimulating a T cell response and a B cell response when expressed in a subject.

[0009] In some aspects, the infectious disease organism is selected from the group consisting of: Severe acute respiratory syndrome-related coronavirus (SARS), severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), Ebola, HIV, Hepatitis B virus (HBV), influenza, Hepatitis C virus (HCV), Human papillomavirus (HPV), Cytomegalovirus (CMV), Chikungunya virus, Respiratory syncytial virus (RSV), Dengue virus, a orthomyxoviridae family virus, and tuberculosis.

[0010] In some aspects, an ordered sequence of each element of the cassette in the composition for delivery of the self-replicating alphavirus-based expression system is described in the formula, from 5' to 3', comprising $P_a-(L5_b-N_c-L3_d)X-(G5_e-U_f)Y-G3_g$ wherein P comprises the second promoter nucleotide sequence, where $a = 0$ or 1 , N comprises one of the epitope-encoding nucleic acid sequences, wherein the epitope-encoding nucleic acid sequence comprises an MHC class I epitope-encoding nucleic acid sequence, where $c = 1$, L5 comprises the 5' linker sequence, where $b = 0$ or 1 , L3 comprises the 3' linker sequence, where $d = 0$ or 1 , G5 comprises one of the at least one nucleic acid sequences encoding a GPGPG amino acid linker, where $e = 0$ or 1 , G3 comprises one of the at least one nucleic acid sequences encoding

a GPGPG amino acid linker, where $g = 0$ or 1 , U comprises one of the at least one MHC class II epitope-encoding nucleic acid sequence, where $f = 1$, $X = 1$ to 400 , where for each X the corresponding N_c is an MHC class I epitope-encoding nucleic acid sequence, and $Y = 0, 1$, or 2 , where for each Y the corresponding U_f is an MHC class II epitope-encoding nucleic acid sequence. In some aspects, for each X the corresponding N_c is a distinct MHC class I epitope-encoding nucleic acid sequence. In some aspects, for each Y the corresponding U_f is a distinct MHC class II epitope-encoding nucleic acid sequence. In some aspects, wherein $a = 0$, $b = 1$, $d = 1$, $e = 1$, $g = 1$, $h = 1$, $X = 20$, $Y = 2$, the at least one promoter nucleotide sequence is a single 26S promoter nucleotide sequence provided by the RNA alphavirus backbone, the at least one polyadenylation poly(A) sequence is a poly(A) sequence of at least 100 consecutive A nucleotides provided by the RNA alphavirus backbone, the cassette is integrated between the 26S promoter nucleotide sequence and the poly(A) sequence, wherein the cassette is operably linked to the 26S promoter nucleotide sequence and the poly(A) sequence, each N encodes a MHC class I epitope 7-15 amino acids in length, $L5$ is a native 5' linker sequence that encodes a native N-terminal amino acid sequence of the MHC I epitope, and wherein the 5' linker sequence encodes a peptide that is at least 3 amino acids in length, $L3$ is a native 3' linker sequence that encodes a native C-terminal amino acid sequence of the MHC I epitope, and wherein the 3' linker sequence encodes a peptide that is at least 3 amino acids in length, U is each of a PADRE class II sequence and a Tetanus toxoid MHC class II sequence, the RNA alphavirus backbone is the sequence set forth in SEQ ID NO:6, and each of the MHC class I epitope-encoding nucleic acid sequences encodes a polypeptide that is between 13 and 25 amino acids in length.

[0011] In some aspects, the LNP comprises a lipid selected from the group consisting of: an ionizable amino lipid, a phosphatidylcholine, cholesterol, a PEG-based coat lipid, or a combination thereof. In some aspects, the LNP comprises an ionizable amino lipid, a phosphatidylcholine, cholesterol, and a PEG-based coat lipid. In some aspects, the ionizable amino lipids comprise MC3-like (dilinoleylmethyl-4-dimethylaminobutyrate) molecules. In some aspects, the LNP-encapsulated expression system has a diameter of about 100nm.

[0012] In some aspects, the composition for delivery of the self-replicating alphavirus-based expression system is formulated for intramuscular (IM), intradermal (ID), subcutaneous (SC), or intravenous (IV) administration. In some aspects, the composition for delivery of the self-replicating alphavirus-based expression system is formulated for intramuscular (IM) administration.

[0013] In some aspects, the cassette is integrated between the at least one promoter nucleotide sequence and the at least one poly(A) sequence. In some aspects, the at least one promoter nucleotide sequence is operably linked to the cassette.

[0014] In some aspects, the one or more vectors comprise one or more +-stranded RNA vectors. In some aspects, the one or more +-stranded RNA vectors comprise a 5' 7-methylguanosine (m7g) cap. In some aspects, the one or more +-stranded RNA vectors are produced by *in vitro* transcription. In some aspects, the one or more vectors are self-replicating within a mammalian cell. In some aspects, the RNA alphavirus backbone comprises at least one nucleotide sequence of an Aura virus, a Fort Morgan virus, a Venezuelan equine encephalitis virus, a Ross River virus, a Semliki Forest virus, a Sindbis virus, or a Mayaro virus. In some aspects, the RNA alphavirus backbone comprises at least one nucleotide sequence of a Venezuelan equine encephalitis virus. In some aspects, the RNA alphavirus backbone comprises at least sequences for nonstructural protein-mediated amplification, a 26S promoter sequence, a poly(A) sequence, a nonstructural protein 1 (nsP1) gene, a nsP2 gene, a nsP3 gene, and a nsP4 gene encoded by the nucleotide sequence of the Aura virus, the Fort Morgan virus, the Venezuelan equine encephalitis virus, the Ross River virus, the Semliki Forest virus, the Sindbis virus, or the Mayaro virus. In some aspects, the RNA alphavirus backbone comprises at least sequences for nonstructural protein-mediated amplification, a 26S promoter sequence, and a poly(A) sequence encoded by the nucleotide sequence of the Aura virus, the Fort Morgan virus, the Venezuelan equine encephalitis virus, the Ross River virus, the Semliki Forest virus, the Sindbis virus, or the Mayaro virus. In some aspects, sequences for nonstructural protein-mediated amplification are selected from the group consisting of: an alphavirus 5' UTR, a 51-nt CSE, a 24-nt CSE, a 26S subgenomic promoter sequence, a 19-nt CSE, an alphavirus 3' UTR, or combinations thereof. In some aspects, the RNA alphavirus backbone does not encode structural virion proteins capsid, E2 and E1. In some aspects, the cassette is inserted in place of structural virion proteins within the nucleotide sequence of the Aura virus, the Fort Morgan virus, the Venezuelan equine encephalitis virus, the Ross River virus, the Semliki Forest virus, the Sindbis virus, or the Mayaro virus. In some aspects, the Venezuelan equine encephalitis virus comprises the sequence of SEQ ID NO:3 or SEQ ID NO:5. In some aspects, the Venezuelan equine encephalitis virus comprises the sequence of SEQ ID NO:3 or SEQ ID NO:5 further comprising a deletion between base pair 7544 and 11175. In some aspects, the RNA alphavirus backbone comprises the sequence set forth in SEQ ID NO:6 or SEQ ID NO:7. In some aspects, the cassette is inserted at position 7544 to replace

the deletion between base pairs 7544 and 11175 as set forth in the sequence of SEQ ID NO:3 or SEQ ID NO:5. In some aspects, the insertion of the cassette provides for transcription of a polycistronic RNA comprising the nsP1-4 genes and the at least one nucleic acid sequence, wherein the nsP1-4 genes and the at least one nucleic acid sequence are in separate open reading frames.

[0015] In some aspects, the at least one promoter nucleotide sequence is the native 26S promoter nucleotide sequence encoded by the RNA alphavirus backbone. In some aspects, the at least one promoter nucleotide sequence is an exogenous RNA promoter. In some aspects, the second promoter nucleotide sequence is a 26S promoter nucleotide sequence. In some aspects, the second promoter nucleotide sequence comprises multiple 26S promoter nucleotide sequences, wherein each 26S promoter nucleotide sequence provides for transcription of one or more of the separate open reading frames.

[0016] In some aspects, the one or more vectors are each at least 300nt in size. In some aspects, the one or more vectors are each at least 1kb in size. In some aspects, the one or more vectors are each 2kb in size. In some aspects, the one or more vectors are each less than 5kb in size.

[0017] In some aspects, the at least one antigen-encoding nucleic acid sequence comprises two or more antigen-encoding nucleic acid sequences. In some aspects, each antigen-encoding nucleic acid sequence is linked directly to one another. In some aspects, each antigen-encoding nucleic acid sequence is linked to a distinct antigen-encoding nucleic acid sequence with a nucleic acid sequence encoding a linker. In some aspects, the linker links two MHC class I epitope-encoding nucleic acid sequences or an MHC class I epitope-encoding nucleic acid sequence to an MHC class II epitope-encoding nucleic acid sequence. In some aspects, the linker is selected from the group consisting of: (1) consecutive glycine residues, at least 2, 3, 4, 5, 6, 7, 8, 9, or 10 residues in length; (2) consecutive alanine residues, at least 2, 3, 4, 5, 6, 7, 8, 9, or 10 residues in length; (3) two arginine residues (RR); (4) alanine, alanine, tyrosine (AAY); (5) a consensus sequence at least 2, 3, 4, 5, 6, 7, 8, 9, or 10 amino acid residues in length that is processed efficiently by a mammalian proteasome; and (6) one or more native sequences flanking the antigen derived from the cognate protein of origin and that is at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, or 2-20 amino acid residues in length. In some aspects, the linker links two MHC class II epitope-encoding nucleic acid sequences or an MHC class II sequence to an MHC class I epitope-encoding nucleic acid sequence. In some aspects, the linker comprises the sequence GPGPG.

[0018] In some aspects, the antigen-encoding nucleic acid sequences is linked, operably or directly, to a separate or contiguous sequence that enhances the expression, stability, cell trafficking, processing and presentation, and/or immunogenicity of the antigen-encoding nucleic acid sequence. In some aspects, the separate or contiguous sequence comprises at least one of: a ubiquitin sequence, a ubiquitin sequence modified to increase proteasome targeting (e.g., the ubiquitin sequence contains a Gly to Ala substitution at position 76), an immunoglobulin signal sequence (e.g., IgK), a major histocompatibility class I sequence, lysosomal-associated membrane protein (LAMP)-1, human dendritic cell lysosomal-associated membrane protein, and a major histocompatibility class II sequence; optionally wherein the ubiquitin sequence modified to increase proteasome targeting is A76.

[0019] In some aspects, the at least one antigen-encoding nucleic acid sequence comprises at least 2-10, 2, 3, 4, 5, 6, 7, 8, 9, or 10 antigen-encoding nucleic acid sequences, optionally wherein each antigen-encoding nucleic acid sequence encodes a distinct antigen-encoding nucleic acid sequence. In some aspects, the at least one antigen-encoding nucleic acid sequence comprises at least 11-20, 15-20, 11-100, 11-200, 11-300, 11-400, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 or up to 400 antigen-encoding nucleic acid sequences, optionally wherein each antigen-encoding nucleic acid sequence encodes a distinct antigen-encoding nucleic acid sequence. In some aspects, the at least one antigen-encoding nucleic acid sequence comprises at least 11-20, 15-20, 11-100, 11-200, 11-300, 11-400, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 or up to 400 antigen-encoding nucleic acid sequences. In some aspects, the at least one antigen-encoding nucleic acid sequence comprises at least 2-400 antigen-encoding nucleic acid sequences and wherein at least two of the antigen-encoding nucleic acid sequences encode epitope sequences or portions thereof that are presented by MHC class I on a cell surface. In some aspects, the MHC class I epitopes are presented by MHC class I on the infected cell surface.

[0020] In some aspects, the epitope-encoding nucleic acid sequences comprises at least one MHC class I epitope-encoding nucleic acid sequence, and wherein each antigen-encoding nucleic acid sequence encodes a polypeptide sequence between 8 and 35 amino acids in length, optionally 9-17, 9-25, 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 or 35 amino acids in length.

[0021] In some aspects, the at least one MHC class II epitope-encoding nucleic acid sequence is present. In some aspects, the at least one MHC class II epitope-encoding nucleic acid sequence is present and comprises at least one MHC class II epitope-encoding nucleic acid sequence that comprises at least one alteration that makes the encoded epitope sequence

distinct from the corresponding peptide sequence encoded by a wild-type nucleic acid sequence. In some aspects, the epitope-encoding nucleic acid sequence comprises an MHC class II epitope-encoding nucleic acid sequence and wherein each antigen-encoding nucleic acid sequence encodes a polypeptide sequence that is 12-20, 12, 13, 14, 15, 16, 17, 18, 19, 20, or 20-40 amino acids in length. In some aspects, the epitope-encoding nucleic acid sequences comprises an MHC class II epitope-encoding nucleic acid sequence, wherein the at least one MHC class II epitope-encoding nucleic acid sequence is present, and wherein the at least one MHC class II epitope-encoding nucleic acid sequence comprises at least one universal MHC class II epitope-encoding nucleic acid sequence, optionally wherein the at least one universal sequence comprises at least one of Tetanus toxoid and PADRE.

[0022] In some aspects, the at least one promoter nucleotide sequence or the second promoter nucleotide sequence is inducible. In some aspects, the at least one promoter nucleotide sequence or the second promoter nucleotide sequence is non-inducible.

[0023] In some aspects, the at least one poly(A) sequence comprises a poly(A) sequence native to the alphavirus. In some aspects, the at least one poly(A) sequence comprises a poly(A) sequence exogenous to the alphavirus. In some aspects, the at least one poly(A) sequence is operably linked to at least one of the at least one nucleic acid sequences. In some aspects, the at least one poly(A) sequence is at least 20, at least 30, at least 40, at least 50, at least 60, at least 70, at least 80, or at least 90 consecutive A nucleotides. In some aspects, the at least one poly(A) sequence is at least 100 consecutive A nucleotides.

[0024] In some aspects, the epitope-encoding nucleic acid sequence comprises a MHC class I epitope-encoding nucleic acid sequence, and wherein the MHC class I epitope-encoding nucleic acid sequence is selected by performing the steps of: (a) obtaining at least one of exome, transcriptome, or whole genome nucleotide sequencing data from the infectious disease organism, wherein the infectious disease organism nucleotide sequencing data is used to obtain data representing peptide sequences of each of a set of epitopes; (b) inputting the peptide sequence of each epitope into a presentation model to generate a set of numerical likelihoods that each of the epitopes is presented by one or more of the MHC alleles on the infected cell surface, the set of numerical likelihoods having been identified at least based on received mass spectrometry data; and (c) selecting a subset of the set of epitopes based on the set of numerical likelihoods to generate a set of selected epitopes which are used to generate the MHC class I epitope-encoding nucleic acid sequence. In some aspects, each of the MHC class I epitope-encoding nucleic acid sequences is selected by performing the steps of: (a) obtaining at least

one of exome, transcriptome, or whole genome infectious disease organism nucleotide sequencing data from the infectious disease organism, wherein the infectious disease organism nucleotide sequencing data is used to obtain data representing peptide sequences of each of a set of epitopes; (b) inputting the peptide sequence of each epitope into a presentation model to generate a set of numerical likelihoods that each of the epitopes is presented by one or more of the MHC alleles on the infected cell surface, the set of numerical likelihoods having been identified at least based on received mass spectrometry data; and; and (c) selecting a subset of the set of epitopes based on the set of numerical likelihoods to generate a set of selected epitopes which are used to generate the at least 20 MHC class I epitope-encoding nucleic acid sequences. In some aspects, a number of the set of selected epitopes is 2-20. In some aspects, the presentation model represents dependence between: (a) presence of a pair of a particular one of the MHC alleles and a particular amino acid at a particular position of a peptide sequence; and (b) likelihood of presentation on the infected cell surface, by the particular one of the MHC alleles of the pair, of such a peptide sequence comprising the particular amino acid at the particular position. In some aspects, selecting the set of selected epitopes comprises selecting epitopes that have an increased likelihood of being presented on the infected cell surface relative to unselected epitopes based on the presentation model. In some aspects, selecting the set of selected epitopes comprises selecting epitopes that have an increased likelihood of being capable of inducing a infectious disease organism-specific immune response in the subject relative to unselected epitopes based on the presentation model. In some aspects, selecting the set of selected epitopes comprises selecting epitopes that have an increased likelihood of being capable of being presented to naïve T cells by professional antigen presenting cells (APCs) relative to unselected epitopes based on the presentation model, optionally wherein the APC is a dendritic cell (DC). In some aspects, selecting the set of selected epitopes comprises selecting epitopes that have a decreased likelihood of being subject to inhibition via central or peripheral tolerance relative to unselected epitopes based on the presentation model. In some aspects, selecting the set of selected epitopes comprises selecting epitopes that have a decreased likelihood of being capable of inducing an autoimmune response to normal tissue in the subject relative to unselected epitopes based on the presentation model. In some aspects, exome or transcriptome nucleotide sequencing data is obtained by performing sequencing on the infected tissue. In some aspects, the sequencing is next generation sequencing (NGS) or any massively parallel sequencing approach.

[0025] Also disclosed herein is a composition for delivery of a chimpanzee adenovirus (ChAdV)-based expression system, wherein the composition for delivery of the ChAdV-based expression system comprises: the ChAdV-based expression system, wherein the ChAdV-based expression system comprises a viral particle comprising a ChAdV vector, wherein the ChAdV vector comprises: (a) a ChAdV backbone, wherein the ChAdV backbone comprises: (i) at least one promoter nucleotide sequence, and (ii) at least one polyadenylation (poly(A)) sequence; and (b) a cassette, wherein the cassette comprises: (i) at least one antigen-encoding nucleic acid sequence comprising: a. a nucleic acid sequence encoding an infectious disease organism peptide selected from the group consisting of: a pathogen-derived peptide, a virus-derived peptide, a bacteria-derived peptide, a fungus-derived peptide, and a parasite-derived peptide b. optionally a 5' linker sequence, and c. optionally a 3' linker sequence; and wherein the cassette is operably linked to the at least one promoter nucleotide sequence and the at least one poly(A) sequence.

[0026] Also disclosed herein is a composition for delivery of a ChAdV-based expression system, wherein the composition for delivery of the ChAdV-based expression system comprises: the ChAdV-based expression system, wherein the ChAdV-based expression system comprises a viral particle comprising a ChAdV vector, wherein the ChAdV vector comprises: (a) a ChAdV backbone, wherein the ChAdV backbone comprises: (i) a modified ChAdV68 sequence comprising at least nucleotides 2 to 36,518 of the sequence set forth in SEQ ID NO:1, wherein the nucleotides 2 to 36,518 lack: (1) nucleotides 577 to 3403 of the sequence shown in SEQ ID NO:1 corresponding to an E1 deletion; (2) nucleotides 27,125 to 31,825 of the sequence shown in SEQ ID NO:1 corresponding to an E3 deletion; and optionally (3) nucleotides 34,916 to 35,642 of the sequence shown in SEQ ID NO:1 corresponding to a partial E4 deletion; (ii) a CMV promoter nucleotide sequence; and (iii) an SV40 polyadenylation (poly(A)) sequence; and (b) a cassette, wherein the cassette comprises: (i) at least one antigen-encoding nucleic acid sequence comprising: a. a nucleic acid sequence encoding an infectious disease organism peptide selected from the group consisting of: a pathogen-derived peptide, a virus-derived peptide, a bacteria-derived peptide, a fungus-derived peptide, and a parasite-derived peptide b. optionally a 5' linker sequence, and c. optionally a 3' linker sequence; and wherein the cassette is inserted within the E1 deletion and the cassette is operably linked to the CMV promoter nucleotide sequence and the SV40 poly(A) sequence.

[0027] In some aspects, the nucleic acid sequence encoding the peptide comprises an epitope-encoding nucleic acid sequence. In some aspects, the nucleic acid sequence encoding

the peptide comprises two or more distinct epitope-encoding nucleic acid sequences. In some aspects, the nucleic acid sequence encoding the peptide comprises two or more distinct epitope-encoding nucleic acid sequences. In some aspects, the nucleic acid sequence encoding the peptide comprises between 1-10, between 1-20, between 1-30, between 1-40, between 1-50, between 1-100, between 1-200, between 1-300, between 1-400, or between 1-500 distinct epitope-encoding nucleic acid sequences distinct epitope-encoding nucleic acid sequences. In some aspects, the nucleic acid sequence encoding the peptide comprises between 2-10, between 2-20, between 2-30, between 2-40, between 2-50, between 2-100, between 2-200, between 2-300, between 2-400, or between 2-500 distinct epitope-encoding nucleic acid sequences distinct epitope-encoding nucleic acid sequences. In some aspects, the nucleic acid sequence encoding the peptide comprises a nucleic acid sequence encoding a full-length protein expressed in the infectious disease organism. In some aspects, the nucleic acid sequence encoding the peptide comprises a nucleic acid sequence encoding a protein domain of the protein expressed in the infectious disease organism. In some aspects, the nucleic acid sequence encoding the peptide comprises a nucleic acid sequence encoding a protein subunit of the protein expressed in the infectious disease organism. In some aspects, the nucleic acid sequence encoding the peptide comprises two or more distinct nucleic acid sequences encoding a peptide selected from: an epitope, a full-length protein, a protein subunit, a protein domain, and combinations thereof of the protein expressed in the infectious disease organism.

[0028] In some aspects, the encoded peptide or peptides is capable of stimulating an immune response when expressed in a subject. In some aspects, the encoded peptide or peptides is capable of stimulating a T cell response when expressed in a subject. In some aspects, the encoded peptide or peptides is capable of stimulating a B cell response when expressed in a subject. In some aspects, the encoded peptide or peptides is capable of stimulating a T cell response and a B cell response when expressed in a subject.

[0029] In some aspects, the infectious disease organism is selected from the group consisting of: Severe acute respiratory syndrome-related coronavirus (SARS), severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), Ebola, HIV, Hepatitis B virus (HBV), influenza, Hepatitis C virus (HCV), Human papillomavirus (HPV), Cytomegalovirus (CMV), Chikungunya virus, Respiratory syncytial virus (RSV), Dengue virus, a orthomyxoviridae family virus, and tuberculosis.

[0030] In some aspects, the epitope-encoding nucleic acid sequence encodes an epitope known or suspected to be presented by MHC class I on a surface of a cell, optionally wherein

the surface of the cell is an infected cell surface, and optionally wherein the cell is a subject's cell. In some aspects, the cell is an infected cell selected from the group consisting of: a pathogen infected cell, a virally infected cell, a bacterially infected cell, a fungally infected cell, and a parasitically infected cell. In some aspects, the virally infected cell is selected from the group consisting of: an HIV infected cell, an HPV infected cell, a SARS infected cell, a SARS-CoV-2 infected cell, an Ebola infected cell, a HBV infected cell, an influenza infected cell, a HCV infected cell, a CMV infected cell, a Chikungunya virus infected cell, a RSV infected cell, a Dengue virus infected cell, a orthomyxoviridae family virus infected cell, and a tuberculosis infected cell.

[0031] In some aspects, an ordered sequence of each element of the cassette in the composition for delivery of the ChAdV-based expression system is described in the formula, from 5' to 3', comprising $P_a-(L5_b-N_c-L3_d)_X-(G5_e-U_f)_Y-G3_g$ wherein P comprises the at least one promoter sequence operably linked to at least one of the at least one antigen-encoding nucleic acid sequences, where $a = 1$, N comprises one of the epitope-encoding nucleic acid sequences, wherein the epitope-encoding nucleic acid sequence comprises an MHC class I epitope-encoding nucleic acid sequence, where $c = 1$, L5 comprises the 5' linker sequence, where $b = 0$ or 1 , L3 comprises the 3' linker sequence, where $d = 0$ or 1 , G5 comprises one of the at least one nucleic acid sequences encoding a GP GPG amino acid linker, where $e = 0$ or 1 , G3 comprises one of the at least one nucleic acid sequences encoding a GP GPG amino acid linker, where $g = 0$ or 1 , U comprises one of the at least one MHC class II epitope-encoding nucleic acid sequence, where $f = 1$, $X = 1$ to 400 , where for each X the corresponding N_c is an MHC class I epitope-encoding nucleic acid sequence, and $Y = 0, 1$, or 2 , where for each Y the corresponding U_f is an MHC class II epitope-encoding nucleic acid sequence. In some aspects, for each X the corresponding N_c is a distinct MHC class I epitope-encoding nucleic acid sequence. In some aspects, for each Y the corresponding U_f is a distinct MHC class II epitope-encoding nucleic acid sequence. In some aspects, $b = 1$, $d = 1$, $e = 1$, $g = 1$, $h = 1$, $X = 10$, $Y = 2$, P is a CMV promoter sequence, each N encodes a MHC class I epitope 7-15 amino acids in length, L5 is a native 5' linker sequence that encodes a native N-terminal amino acid sequence of the MHC I epitope, and wherein the 5' linker sequence encodes a peptide that is at least 3 amino acids in length, L3 is a native 3' linker sequence that encodes a native C-terminal amino acid sequence of the MHC I epitope, and wherein the 3' linker sequence encodes a peptide that is at least 3 amino acids in length, U is each of a PADRE class II sequence and a Tetanus toxoid MHC class II sequence, the ChAdV vector comprises a modified ChAdV68 sequence

comprising at least nucleotides 2 to 36,518 of the sequence set forth in SEQ ID NO:1, wherein the nucleotides 2 to 36,518 lack: (1) nucleotides 577 to 3403 of the sequence shown in SEQ ID NO:1 corresponding to an E1 deletion; (2) nucleotides 27,125 to 31,825 of the sequence shown in SEQ ID NO:1 corresponding to an E3 deletion; and optionally (3) nucleotides 34,916 to 35,642 of the sequence shown in SEQ ID NO:1 corresponding to a partial E4 deletion, and the antigen cassette is inserted within the E1 deletion, and each of the MHC class I antigen-encoding nucleic acid sequences encodes a polypeptide that is 25 amino acids in length.

[0032] In some aspects, the composition for delivery of the ChAdV-based expression system is formulated for intramuscular (IM), intradermal (ID), subcutaneous (SC), or intravenous (IV) administration. In some aspects, the composition for delivery of the ChAdV-based expression system is formulated for intramuscular (IM) administration.

[0033] In some aspects, the cassette is integrated between the at least one promoter nucleotide sequence and the at least one poly(A) sequence. In some aspects, the at least one promoter nucleotide sequence is operably linked to the cassette.

[0034] In some aspects, the ChAdV backbone comprises a ChAdV68 vector backbone. In some aspects, the ChAdV68 vector backbone comprises the sequence set forth in SEQ ID NO:1. In some aspects, the ChAdV68 vector backbone comprises a functional deletion in at least one gene selected from the group consisting of an adenovirus E1A, E1B, E2A, E2B, E3, L1, L2, L3, L4, and L5 gene with reference to a ChAdV68 genome or with reference to the sequence shown in SEQ ID NO:1, optionally wherein the adenoviral backbone or modified ChAdV68 sequence is fully deleted or functionally deleted in: (1) E1A and E1B; or (2) E1A, E1B, and E3 with reference to the adenovirus genome or with reference to the sequence shown in SEQ ID NO:1, optionally wherein the E1 gene is functionally deleted through an E1 deletion of at least nucleotides 577 to 3403 with reference to the sequence shown in SEQ ID NO:1 and optionally wherein the E3 gene is functionally deleted through an E3 deletion of at least nucleotides 27,125 to 31,825 with reference to the sequence shown in SEQ ID NO:1. In some aspects, the ChAdV68 vector backbone comprises one or more deletions between base pair number 577 and 3403 or between base pair 456 and 3014, and optionally wherein the vector further comprises one or more deletions between base pair 27,125 and 31,825 or between base pair 27,816 and 31,333 of the sequence set forth in SEQ ID NO:1. In some aspects, the ChAdV68 vector backbone comprises at least nucleotides 2 to 36,518 of the sequence set forth in SEQ ID NO:1, wherein the nucleotides 2 to 36,518 lack: A. nucleotides 577 to 3403 of the sequence shown in SEQ ID NO:1 corresponding to an E1 deletion; B. nucleotides 27,125 to

31,825 of the sequence shown in SEQ ID NO:1 corresponding to an E3 deletion; C. nucleotides 34,916 to 35,642 of the sequence shown in SEQ ID NO:1 corresponding to a partial E4 deletion; D. nucleotides 456 to 3014 of the sequence shown in SEQ ID NO:1; E. nucleotides 27,816 to 31,333 of the sequence shown in SEQ ID NO:1; F. nucleotides 3957 to 10346 of the sequence shown in SEQ ID NO:1; G. nucleotides 21787 to 23370 of the sequence shown in SEQ ID NO:1; H. nucleotides 33486 to 36193 of the sequence shown in SEQ ID NO:1; or combinations thereof.

[0035] In some aspects, the ChAdV68 vector backbone comprises a partially deleted E4 gene. In some aspects, the partially deleted E4 gene comprises: A. the E4 gene sequence shown in SEQ ID NO:1 and that lacks at least nucleotides 34,916 to 35,642 of the sequence shown in SEQ ID NO:1, B. the E4 gene sequence shown in SEQ ID NO:1 and that lacks at least nucleotides 34,916 to 34,942, nucleotides 34,952 to 35,305 of the sequence shown in SEQ ID NO:1, nucleotides 35,302 to 35,642 of the sequence shown in SEQ ID NO:1, and wherein the vector comprises at least nucleotides 2 to 36,518 of the sequence shown in SEQ ID NO:1, C. the E4 gene sequence shown in SEQ ID NO:1 and that lacks at least nucleotides 34,980 to 36,516 of the sequence shown in SEQ ID NO:1, and wherein the vector comprises at least nucleotides 2 to 36,518 of the sequence shown in SEQ ID NO:1, D. the E4 gene sequence shown in SEQ ID NO:1 and that lacks at least nucleotides 34,979 to 35,642 of the sequence shown in SEQ ID NO:1, and wherein the vector comprises at least nucleotides 2 to 36,518 of the sequence shown in SEQ ID NO:1, E. an E4 deletion of at least a partial deletion of E4Orf2, a fully deleted E4Orf3, and at least a partial deletion of E4Orf4, F. an E4 deletion of at least a partial deletion of E4Orf2, at least a partial deletion of E4Orf3, and at least a partial deletion of E4Orf4, G. an E4 deletion of at least a partial deletion of E4Orf1, a fully deleted E4Orf2, and at least a partial deletion of E4Orf3, or H. an E4 deletion of at least a partial deletion of E4Orf2 and at least a partial deletion of E4Orf3.

[0036] In some aspects, the cassette is inserted in the ChAdV backbone at the E1 region, E3 region, and/or any deleted AdV region that allows incorporation of the cassette. In some aspects, the ChAdV backbone is generated from one of a first generation, a second generation, or a helper-dependent adenoviral vector.

[0037] In some aspects, the at least one promoter nucleotide sequence is selected from the group consisting of: a CMV, a SV40, an EF-1, a RSV, a PGK, a HSA, a MCK, and a EBV promoter sequence. In some aspects, the at least one promoter nucleotide sequence is a CMV promoter sequence.

[0038] In some aspects, at least one of the epitope-encoding nucleic acid sequences encodes an epitope that, when expressed and translated, is capable of being presented by MHC class I on a cell of a subject. In some aspects, at least one of the epitope-encoding nucleic acid sequences encodes an epitope that, when expressed and translated, is capable of being presented by MHC class II on a cell of a subject.

[0039] In some aspects, the at least one antigen-encoding nucleic acid sequence comprises two or more antigen-encoding nucleic acid sequences. In some aspects, each antigen-encoding nucleic acid sequence is linked directly to one another. In some aspects, each antigen-encoding nucleic acid sequence is linked to a distinct antigen-encoding nucleic acid sequence with a nucleic acid sequence encoding a linker. In some aspects, the linker links two MHC class I epitope-encoding nucleic acid sequences or an MHC class I epitope-encoding nucleic acid sequence to an MHC class II epitope-encoding nucleic acid sequence. In some aspects, the linker is selected from the group consisting of: (1) consecutive glycine residues, at least 2, 3, 4, 5, 6, 7, 8, 9, or 10 residues in length; (2) consecutive alanine residues, at least 2, 3, 4, 5, 6, 7, 8, 9, or 10 residues in length; (3) two arginine residues (RR); (4) alanine, alanine, tyrosine (AAY); (5) a consensus sequence at least 2, 3, 4, 5, 6, 7, 8, 9, or 10 amino acid residues in length that is processed efficiently by a mammalian proteasome; and (6) one or more native sequences flanking the antigen derived from the cognate protein of origin and that is at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, or 2-20 amino acid residues in length. In some aspects, the linker links two MHC class II epitope-encoding nucleic acid sequences or an MHC class II sequence to an MHC class I epitope-encoding nucleic acid sequence. In some aspects, the linker comprises the sequence GPGPG.

[0040] In some aspects, the antigen-encoding nucleic acid sequences is linked, operably or directly, to a separate or contiguous sequence that enhances the expression, stability, cell trafficking, processing and presentation, and/or immunogenicity of the antigen-encoding nucleic acid sequence. In some aspects, the separate or contiguous sequence comprises at least one of: a ubiquitin sequence, a ubiquitin sequence modified to increase proteasome targeting (e.g., the ubiquitin sequence contains a Gly to Ala substitution at position 76), an immunoglobulin signal sequence (e.g., IgK), a major histocompatibility class I sequence, lysosomal-associated membrane protein (LAMP)-1, human dendritic cell lysosomal-associated membrane protein, and a major histocompatibility class II sequence; optionally wherein the ubiquitin sequence modified to increase proteasome targeting is A76.

[0041] In some aspects, the epitope-encoding nucleic acid sequence comprises at least one alteration that makes the encoded epitope have increased binding affinity to its corresponding MHC allele relative to the translated, corresponding wild-type nucleic acid sequence. In some aspects, the epitope-encoding nucleic acid sequence comprises at least one alteration that makes the encoded epitope have increased binding stability to its corresponding MHC allele relative to the translated, corresponding wild-type nucleic acid sequence. In some aspects, the epitope-encoding nucleic acid sequence comprises at least one alteration that makes the encoded epitope have an increased likelihood of presentation on its corresponding MHC allele relative to the translated, corresponding wild-type nucleic acid sequence. In some aspects, the at least one alteration comprises a point mutation, a frameshift mutation, a non-frameshift mutation, a deletion mutation, an insertion mutation, a splice variant, a genomic rearrangement, or a proteasome-generated spliced antigen. In some aspects, the epitope-encoding nucleic acid sequence encodes an epitope known or suspected to be expressed in a subject known or suspected to have an infection. In some aspects, the infection is selected from the group consisting of: a pathogen infection, a viral infection, a bacterial infection, an fungal infection, and a parasitic infection. In some aspects, the viral infection is selected from the group consisting of: an HIV infection, an HPV infection, a SARS infection, a SARS-CoV-2 infection, an Ebola infection, a HBV infection, an influenza infection, a HCV infection, a CMV infection, a Chikungunya virus infection, a RSV infection, a Dengue virus infection, a orthomyxoviridae family virus infection, and a tuberculosis infection. In some aspects, the bacterial infection is a tuberculosis infection.

[0042] In some aspects, the at least one antigen-encoding nucleic acid sequence comprises at least 2-10, 2, 3, 4, 5, 6, 7, 8, 9, or 10 antigen-encoding nucleic acid sequences, optionally wherein each antigen-encoding nucleic acid sequence encodes a distinct antigen-encoding nucleic acid sequence. In some aspects, the at least one antigen-encoding nucleic acid sequence comprises at least 11-20, 15-20, 11-100, 11-200, 11-300, 11-400, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 or up to 400 antigen-encoding nucleic acid sequences, optionally wherein each antigen-encoding nucleic acid sequence encodes a distinct antigen-encoding nucleic acid sequence. In some aspects, the at least one antigen-encoding nucleic acid sequence comprises at least 11-20, 15-20, 11-100, 11-200, 11-300, 11-400, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 or up to 400 antigen-encoding nucleic acid sequences. In some aspects, the at least one antigen-encoding nucleic acid sequence comprises at least 2-400 antigen-encoding nucleic acid sequences and wherein at least two of the antigen-encoding nucleic acid sequences encode epitope sequences

or portions thereof that are presented by MHC class I on a cell surface. In some aspects, at least two of the MHC class I epitopes are presented by MHC class I on the infected cell surface. In some aspects, the epitope-encoding nucleic acid sequences comprises at least one MHC class I epitope-encoding nucleic acid sequence, and wherein each antigen-encoding nucleic acid sequence encodes a polypeptide sequence between 8 and 35 amino acids in length, optionally 9-17, 9-25, 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 or 35 amino acids in length.

[0043] In some aspects, the at least one MHC class II epitope-encoding nucleic acid sequence is present. In some aspects, the at least one MHC class II epitope-encoding nucleic acid sequence is present and comprises at least one MHC class II epitope-encoding nucleic acid sequence that comprises at least one alteration that makes the encoded epitope sequence distinct from the corresponding peptide sequence encoded by a wild-type nucleic acid sequence. In some aspects, the epitope-encoding nucleic acid sequence comprises an MHC class II epitope-encoding nucleic acid sequence and wherein each antigen-encoding nucleic acid sequence encodes a polypeptide sequence that is 12-20, 12, 13, 14, 15, 16, 17, 18, 19, 20, or 20-40 amino acids in length. In some aspects, the epitope-encoding nucleic acid sequences comprises an MHC class II epitope-encoding nucleic acid sequence, wherein the at least one MHC class II epitope-encoding nucleic acid sequence is present, and wherein the at least one MHC class II epitope-encoding nucleic acid sequence comprises at least one universal MHC class II epitope-encoding nucleic acid sequence, optionally wherein the at least one universal sequence comprises at least one of Tetanus toxoid and PADRE.

[0044] In some aspects, the at least one promoter nucleotide sequence is inducible. In some aspects, the at least one promoter nucleotide sequence is non-inducible.

[0045] In some aspects, the at least one poly(A) sequence comprises a Bovine Growth Hormone (BGH) SV40 polyA sequence. In some aspects, the at least one poly(A) sequence is at least 20, at least 30, at least 40, at least 50, at least 60, at least 70, at least 80, or at least 90 consecutive A nucleotides. In some aspects, the at least one poly(A) sequence is at least 100 consecutive A nucleotides.

[0046] In some aspects, the cassette further comprises at least one of: an intron sequence, a woodchuck hepatitis virus posttranscriptional regulatory element (WPRE) sequence, an internal ribosome entry sequence (IRES) sequence, a nucleotide sequence encoding a 2A self cleaving peptide sequence, a nucleotide sequence encoding a Furin cleavage site, or a sequence in the 5' or 3' non-coding region known to enhance the nuclear export, stability, or translation efficiency

of mRNA that is operably linked to at least one of the at least one antigen-encoding nucleic acid sequences. In some aspects, the cassette further comprises a reporter gene, including but not limited to, green fluorescent protein (GFP), a GFP variant, secreted alkaline phosphatase, luciferase, a luciferase variant, or a detectable peptide or epitope. In some aspects, the detectable peptide or epitope is selected from the group consisting of an HA tag, a Flag tag, a His-tag, or a V5 tag.

[0047] In some aspects, the one or more vectors further comprises one or more nucleic acid sequences encoding at least one immune modulator. In some aspects, the immune modulator is an anti-CTLA4 antibody or an antigen-binding fragment thereof, an anti-PD-1 antibody or an antigen-binding fragment thereof, an anti-PD-L1 antibody or an antigen-binding fragment thereof, an anti-4-1BB antibody or an antigen-binding fragment thereof, or an anti-OX-40 antibody or an antigen-binding fragment thereof. In some aspects, the antibody or antigen-binding fragment thereof is a Fab fragment, a Fab' fragment, a single chain Fv (scFv), a single domain antibody (sdAb) either as single specific or multiple specificities linked together (e.g., camelid antibody domains), or full-length single-chain antibody (e.g., full-length IgG with heavy and light chains linked by a flexible linker). In some aspects, the heavy and light chain sequences of the antibody are a contiguous sequence separated by either a self-cleaving sequence such as 2A or IRES; or the heavy and light chain sequences of the antibody are linked by a flexible linker such as consecutive glycine residues. In some aspects, the immune modulator is a cytokine. In some aspects, the cytokine is at least one of IL-2, IL-7, IL-12, IL-15, or IL-21 or variants thereof of each.

[0048] In some aspects, the epitope-encoding nucleic acid sequence comprises a MHC class I epitope-encoding nucleic acid sequence, and wherein the MHC class I epitope-encoding nucleic acid sequence is selected by performing the steps of: (a) obtaining at least one of exome, transcriptome, or whole genome infectious disease organism nucleotide sequencing data from the infectious disease organism, wherein the infectious disease organism nucleotide sequencing data is used to obtain data representing peptide sequences of each of a set of epitopes; (b) inputting the peptide sequence of each epitope into a presentation model to generate a set of numerical likelihoods that each of the epitopes is presented by one or more of the MHC alleles on the infected cell surface, the set of numerical likelihoods having been identified at least based on received mass spectrometry data; and (c) selecting a subset of the set of epitopes based on the set of numerical likelihoods to generate a set of selected epitopes which are used to generate the MHC class I epitope-encoding nucleic acid sequence. In some

aspects, each of the MHC class I epitope-encoding nucleic acid sequences is selected by performing the steps of: (a) obtaining at least one of exome, transcriptome, or whole genome infectious disease organism nucleotide sequencing data from the infectious disease organism, wherein the infectious disease organism nucleotide sequencing data is used to obtain data representing peptide sequences of each of a set of epitopes; (b) inputting the peptide sequence of each epitope into a presentation model to generate a set of numerical likelihoods that each of the epitopes is presented by one or more of the MHC alleles on the infected cell surface, the set of numerical likelihoods having been identified at least based on received mass spectrometry data; and (c) selecting a subset of the set of epitopes based on the set of numerical likelihoods to generate a set of selected epitopes which are used to generate the at least 20 MHC class I epitope-encoding nucleic acid sequences. In some aspects, a number of the set of selected epitopes is 2-20. In some aspects, the presentation model represents dependence between: (a) presence of a pair of a particular one of the MHC alleles and a particular amino acid at a particular position of a peptide sequence; and (b) likelihood of presentation on the infected cell surface, by the particular one of the MHC alleles of the pair, of such a peptide sequence comprising the particular amino acid at the particular position. In some aspects, selecting the set of selected epitopes comprises selecting epitopes that have an increased likelihood of being presented on the infected cell surface relative to unselected epitopes based on the presentation model. In some aspects, selecting the set of selected epitopes comprises selecting epitopes that have an increased likelihood of being capable of inducing a infectious disease organism-specific immune response in the subject relative to unselected epitopes based on the presentation model. In some aspects, selecting the set of selected epitopes comprises selecting epitopes that have an increased likelihood of being capable of being presented to naïve T cells by professional antigen presenting cells (APCs) relative to unselected epitopes based on the presentation model, optionally wherein the APC is a dendritic cell (DC). In some aspects, selecting the set of selected epitopes comprises selecting epitopes that have a decreased likelihood of being subject to inhibition via central or peripheral tolerance relative to unselected epitopes based on the presentation model. In some aspects, selecting the set of selected epitopes comprises selecting epitopes that have a decreased likelihood of being capable of inducing an autoimmune response to normal tissue in the subject relative to unselected epitopes based on the presentation model. In some aspects, exome or transcriptome nucleotide sequencing data is obtained by performing sequencing on the infected tissue. In

some aspects, the sequencing is next generation sequencing (NGS) or any massively parallel sequencing approach.

[0049] In some aspects, the cassette comprises junctional epitope sequences formed by adjacent sequences in the cassette. In some aspects, at least one or each junctional epitope sequence has an affinity of greater than 500 nM for MHC. In some aspects, each junctional epitope sequence is non-self. In some aspects, the cassette does not encode a non-therapeutic MHC class I or class II epitope nucleic acid sequence comprising a translated, wild-type nucleic acid sequence, wherein the non-therapeutic epitope is predicted to be displayed on an MHC allele of the subject. In some aspects, the non-therapeutic predicted MHC class I or class II epitope sequence is a junctional epitope sequence formed by adjacent sequences in the cassette. In some aspects, the prediction is based on presentation likelihoods generated by inputting sequences of the non-therapeutic epitopes into a presentation model. In some aspects, an order of the antigen-encoding nucleic acid sequences in the cassette is determined by a series of steps comprising: (a) generating a set of candidate cassette sequences corresponding to different orders of the antigen-encoding nucleic acid sequences; (b) determining, for each candidate cassette sequence, a presentation score based on presentation of non-therapeutic epitopes in the candidate cassette sequence; and (c) selecting a candidate cassette sequence associated with a presentation score below a predetermined threshold as the cassette sequence for a vaccine.

[0050] In some aspects, the composition for delivery of the ChAdV-based expression system is formulated in a pharmaceutical composition comprising a pharmaceutically acceptable carrier.

[0051] In some aspects, one or more of the epitope-encoding nucleic acid sequences are derived from an infection in or an infected cell of a subject. In some aspects, each of the epitope-encoding nucleic acid sequences are derived from an infection in or an infected cell of a subject. In some aspects, one or more of the epitope-encoding nucleic acid sequences are not derived from an infection in or an infected cell of a subject. In some aspects, each of the epitope-encoding nucleic acid sequences are not derived from an infection in or an infected cell of a subject.

[0052] In some aspects, the at least one antigen-encoding nucleic acid sequence comprises at least 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 infectious disease organism-specific MHC class I antigen-encoding nucleic acid sequences linearly linked to each other. In some aspects, the at least one epitope-encoding nucleic acid sequence comprises at least 10, 11, 12, 13, 14,

15, 16, 17, 18, 19, or 20 infectious disease organism-specific MHC class I epitope-encoding nucleic acid sequences linearly linked to each other.

[0053] Also provided for herein is a kit comprising any of the compositions for delivery of the ChAdV-based expression system described herein, and instructions for use.

[0054] Also provided for herein is a composition for delivery of a self-replicating alphavirus-based expression system, wherein the composition for delivery of the self-replicating alphavirus-based expression system comprises: (A) the self-replicating alphavirus-based expression system, wherein the self-replicating alphavirus-based expression system comprises one or more vectors, wherein the one or more vectors comprises: (a) an RNA alphavirus backbone, wherein the RNA alphavirus backbone comprises: (i) at least one promoter nucleotide sequence, and (ii) at least one polyadenylation (poly(A)) sequence; and (b) a cassette, wherein the cassette comprises: (i) at least one antigen-encoding nucleic acid sequence comprising: a. a nucleic acid sequence encoding an infectious disease organism peptide selected from the group consisting of: a pathogen-derived peptide, a virus-derived peptide, a bacteria-derived peptide, a fungus-derived peptide, and a parasite-derived peptide, wherein the infectious disease organism is selected from the group consisting of: Severe acute respiratory syndrome-related coronavirus (SARS), severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), Ebola, HIV, Hepatitis B virus (HBV), influenza, Hepatitis C virus (HCV), Human papillomavirus (HPV), Cytomegalovirus (CMV), Chikungunya virus, Respiratory syncytial virus (RSV), Dengue virus, a orthomyxoviridae family virus, and tuberculosis; b. optionally a 5' linker sequence, and c. optionally a 3' linker sequence; (ii) optionally, a second promoter nucleotide sequence operably linked to the at least one antigen-encoding nucleic acid sequence; and (iii) optionally, at least one second poly(A) sequence, wherein the second poly(A) sequence is a native poly(A) sequence or an exogenous poly(A) sequence to the alphavirus, and (B) a lipid-nanoparticle (LNP), wherein the LNP encapsulates the self-replicating alphavirus-based expression system.

[0055] Also provided for herein is a composition for delivery of a chimpanzee adenovirus (ChAdV)-based expression system, wherein the composition for delivery of the ChAdV-based expression system comprises: the ChAdV-based expression system, wherein the ChAdV-based expression system comprises a viral particle comprising a ChAdV vector, wherein the ChAdV vector comprises: (a) a ChAdV backbone, wherein the ChAdV backbone comprises: (i) at least one promoter nucleotide sequence, and (ii) at least one polyadenylation (poly(A)) sequence; and (b) a cassette, wherein the cassette comprises: (i) at least one antigen-encoding nucleic acid

sequence comprising: a. a nucleic acid sequence encoding an infectious disease organism peptide selected from the group consisting of: a pathogen-derived peptide, a virus-derived peptide, a bacteria-derived peptide, a fungus-derived peptide, and a parasite-derived peptide, wherein the infectious disease organism is selected from the group consisting of: Severe acute respiratory syndrome-related coronavirus (SARS), severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), Ebola, HIV, Hepatitis B virus (HBV), influenza, Hepatitis C virus (HCV), Human papillomavirus (HPV), Cytomegalovirus (CMV), Chikungunya virus, Respiratory syncytial virus (RSV), Dengue virus, a orthomyxoviridae family virus, and tuberculosis; b. optionally a 5' linker sequence, and c. optionally a 3' linker sequence; and wherein the cassette is operably linked to the at least one promoter nucleotide sequence and the at least one poly(A) sequence.

[0056] In some aspects the self-replicating alphavirus-based expression system or the ChAdV-based expression system comprise any of the features of any one of the above claims.

[0057] In some aspects, the cassette of the composition for delivery of the ChAdV-based expression system is identical to the cassette of the composition for delivery of the self-replicating alphavirus-based expression system.

[0058] Also disclosed herein is method for stimulating an immune response in a subject, the method comprising administering to the subject a composition for delivery of a self-replicating alphavirus-based expression system and/or administering to the subject a composition for delivery of a chimpanzee adenovirus (ChAdV)-based expression system, and wherein either: a. the composition for delivery of the ChAdV-based expression system comprises the ChAdV-based expression system, b. wherein the composition for delivery of the self-replicating alphavirus-based expression system comprises the self-replicating alphavirus-based expression system, or c. the composition for delivery of the ChAdV-based expression system comprises the ChAdV-based expression system, and wherein the composition for delivery of the self-replicating alphavirus-based expression system comprises the self-replicating alphavirus-based expression system.

[0059] In some aspects, the composition for delivery of the ChAdV-based expression system is administered as a priming dose and either the composition for delivery of the ChAdV-based expression system or the composition for delivery of the self-replicating alphavirus-based expression system is administered as one or more boosting doses. In some aspects, the composition for delivery of the self-replicating alphavirus-based expression system is administered as a priming dose either the composition for delivery of the ChAdV-based

expression system or the composition for delivery of the self-replicating alphavirus-based expression system is administered as one or more boosting doses. In some aspects, two or more boosting doses are administered. In some aspects, 1, 2, 3, 4, 5, 6, 7, or 8 boosting doses are administered.

[0060] In some aspects, the composition for delivery of the ChAdV-based expression system is administered intramuscularly (IM), intradermally (ID), subcutaneously (SC), or intravenously (IV). In some aspects, the composition for delivery of the ChAdV-based expression system is administered (IM). In some aspects, the IM administration is administered at separate injection sites. In some aspects, the separate injection sites are in opposing deltoid muscles. In some aspects, the separate injection sites are in gluteus or rectus femoris sites on each side.

[0061] In some aspects, the composition for delivery of the self-replicating alphavirus-based expression system is administered intramuscularly (IM), intradermally (ID), subcutaneously (SC), or intravenously (IV). In some aspects, the composition for delivery of the self-replicating alphavirus-based expression system is administered (IM). In some aspects, the IM administration is administered at separate injection sites. In some aspects, the separate injection sites are in opposing deltoid muscles. In some aspects, the separate injection sites are in gluteus or rectus femoris sites on each side. In some aspects, the injection site of the one or more boosting doses is as close as possible to the injection site of the priming dose.

[0062] In some aspects, the method further comprises determining or having determined the HLA-haplotype of the subject.

[0063] In some aspects, the method further comprises administering nivolumab. In some aspects, nivolumab is administered as an intravenous (IV) infusion. In some aspects, nivolumab is administered at a dose of 480 mg. In some aspects, nivolumab is administered on day 1. In some aspects, nivolumab is administered on day 1 and administered every 4 weeks (Q4W) following the priming dose. In some aspects, nivolumab is administered on the same day as the priming dose or on the same day as the one or more boosting doses. In some aspects, nivolumab is formulated in solution at 10 mg/mL.

[0064] In some aspects, the method further comprises administering ipilimumab. In some aspects, ipilimumab is administered as an intravenous (IV) infusion. In some aspects, ipilimumab is administered subcutaneously (SC). In some aspects, the SC administration is injected proximally (within ~2 cm) to one or more of the priming dose injection site or the one or more boosting dose injection sites. In some aspects, the SC administration is administered as 4

separate injections or administered as 6 separate injections. In some aspects, ipilimumab is administered at a dose of 30 mg. In some aspects, ipilimumab is administered on day 1. In some aspects, ipilimumab is administered on day 1 and administered every 4 weeks (Q4W) following the priming dose. In some aspects, ipilimumab is administered on the same day as the priming dose or on the same day as the one or more boosting doses. In some aspects, ipilimumab is formulated in solution at 5 mg/mL.

[0065] In some aspects, the composition for delivery of the self-replicating alphavirus-based expression system comprises: (A) the self-replicating alphavirus-based expression system, wherein the self-replicating alphavirus-based expression system comprises one or more vectors, wherein the one or more vectors comprises: (a) an RNA alphavirus backbone, wherein the RNA alphavirus backbone comprises: (i) at least one promoter nucleotide sequence, and (ii) at least one polyadenylation (poly(A)) sequence; and (b) a cassette, wherein the cassette comprises: (i) at least one antigen-encoding nucleic acid sequence comprising: a. a nucleic acid sequence encoding an infectious disease organism peptide selected from the group consisting of: a pathogen-derived peptide, a virus-derived peptide, a bacteria-derived peptide, a fungus-derived peptide, and a parasite-derived peptide b. optionally a 5' linker sequence, and c. optionally a 3' linker sequence; (ii) optionally, a second promoter nucleotide sequence operably linked to the at least one antigen-encoding nucleic acid sequence; and (iii) optionally, at least one second poly(A) sequence, wherein the second poly(A) sequence is a native poly(A) sequence or an exogenous poly(A) sequence to the alphavirus, and (B) a lipid-nanoparticle (LNP), wherein the LNP encapsulates the self-replicating alphavirus-based expression system.

[0066] In some aspects, the composition for delivery of the self-replicating alphavirus-based expression system comprises, (A) the self-replicating alphavirus-based expression system, wherein the self-replicating alphavirus-based expression system comprises one or more vectors, wherein the one or more vectors comprises: (a) an RNA alphavirus backbone, wherein the RNA alphavirus backbone comprises the nucleic acid sequence set forth in SEQ ID NO:6, wherein the RNA alphavirus backbone sequence comprises a 26S promoter nucleotide sequence and a poly(A) sequence, wherein the 26S promoter sequence is endogenous to the RNA alphavirus backbone, and wherein the poly(A) sequence is endogenous to the RNA alphavirus backbone; and (b) a cassette integrated between the 26S promoter nucleotide sequence and the poly(A) sequence, wherein the cassette is operably linked to the 26S promoter nucleotide sequence, and wherein the cassette comprises at least one antigen-encoding nucleic acid sequence comprising: a. a nucleic acid sequence encoding an infectious

disease organism peptide selected from the group consisting of: a pathogen-derived peptide, a virus-derived peptide, a bacteria-derived peptide, a fungus-derived peptide, and a parasite-derived peptide b. optionally a 5' linker sequence, and c. optionally a 3' linker sequence; and (B) a lipid-nanoparticle (LNP), wherein the LNP encapsulates the self-replicating alphavirus-based expression system.

[0067] In some aspects, the composition for delivery of the self-replicating alphavirus-based expression system comprises at least 30µg of each of the one or more vectors. In some aspects, the composition for delivery of the self-replicating alphavirus-based expression system comprises at least 100µg of each of the one or more vectors. In some aspects, the composition for delivery of the self-replicating alphavirus-based expression system comprises at least 300µg of each of the one or more vectors. In some aspects, the composition for delivery of the self-replicating alphavirus-based expression system comprises at least 400µg, at least 500µg, at least 600µg, at least 700µg, at least 800µg, at least 900µg, at least 1000µg of each of the one or more vectors. In some aspects, the composition for delivery of the self-replicating alphavirus-based expression system comprises between 10-30µg, 10-100µg, 10-300µg, 30-100µg, 30-300µg, or 100-300µg of each of the one or more vectors. In some aspects, the composition for delivery of the self-replicating alphavirus-based expression system comprises between 10-500µg, 10-1000µg, 30-500µg, 30-1000µg, or 500-1000µg of each of the one or more vectors. In some aspects, the composition for delivery of the self-replicating alphavirus-based expression system comprises 400µg, 500µg, 600µg, 700µg, 800µg, 900µg, or 1000µg of each of the one or more vectors. In some aspects, the composition for delivery of the self-replicating alphavirus-based expression system comprises 10 µg, 30µg, 100µg, or 300µg of each of the one or more vectors. In some aspects, the composition for delivery of the self-replicating alphavirus-based expression system comprises less than or equal to 300µg of each of the one or more vectors.

[0068] In some aspects, the weight to weight ratio of the LNP to total weight of the one or more vectors is between 10-40 to 1. In some aspects, the weight to weight ratio of the LNP to total weight of the one or more vectors is between 16-32 to 1. In some aspects, the weight to weight ratio of the LNP to total weight of the one or more vectors is about 24 to 1. In some aspects, the weight to weight ratio of the LNP to total weight of the one or more vectors is 24 to 1.

[0069] In some aspects, the one or more vectors is at a concentration of 1 mg/mL.

[0070] In some aspects, an ordered sequence of each element of the cassette in the composition for delivery of the self-replicating alphavirus-based expression system is described in the formula, from 5' to 3', comprising $P_a-(L5_b-N_c-L3_d)_X-(G5_e-U_f)_Y-G3_g$ wherein P comprises the second promoter nucleotide sequence, where $a = 0$ or 1 , N comprises one of the epitope-encoding nucleic acid sequences, wherein the epitope-encoding nucleic acid sequence comprises an MHC class I epitope-encoding nucleic acid sequence, where $c = 1$, L5 comprises the 5' linker sequence, where $b = 0$ or 1 , L3 comprises the 3' linker sequence, where $d = 0$ or 1 , G5 comprises one of the at least one nucleic acid sequences encoding a GP GPG amino acid linker, where $e = 0$ or 1 , G3 comprises one of the at least one nucleic acid sequences encoding a GP GPG amino acid linker, where $g = 0$ or 1 , U comprises one of the at least one MHC class II epitope-encoding nucleic acid sequence, where $f = 1$, $X = 1$ to 400 , where for each X the corresponding N_c is an MHC class I epitope-encoding nucleic acid sequence, and $Y = 0, 1$, or 2 , where for each Y the corresponding U_f is an MHC class II epitope-encoding nucleic acid sequence. In some aspects, for each X the corresponding N_c is a distinct MHC class I epitope-encoding nucleic acid sequence. In some aspects, for each Y the corresponding U_f is a distinct MHC class II epitope-encoding nucleic acid sequence. In some aspects, $a = 0$, $b = 1$, $d = 1$, $e = 1$, $g = 1$, $h = 1$, $X = 20$, $Y = 2$, the at least one promoter nucleotide sequence is a single 26S promoter nucleotide sequence provided by the RNA alphavirus backbone, the at least one polyadenylation poly(A) sequence is a poly(A) sequence of at least 100 consecutive A nucleotides provided by the RNA alphavirus backbone, the cassette is integrated between the 26S promoter nucleotide sequence and the poly(A) sequence, wherein the cassette is operably linked to the 26S promoter nucleotide sequence and the poly(A) sequence, each N encodes a MHC class I epitope 7-15 amino acids in length, L5 is a native 5' linker sequence that encodes a native N-terminal amino acid sequence of the MHC I epitope, and wherein the 5' linker sequence encodes a peptide that is at least 3 amino acids in length, L3 is a native 3' linker sequence that encodes a native C-terminal amino acid sequence of the MHC I epitope, and wherein the 3' linker sequence encodes a peptide that is at least 3 amino acids in length, U is each of a PADRE class II sequence and a Tetanus toxoid MHC class II sequence, the RNA alphavirus backbone is the sequence set forth in SEQ ID NO:6, and each of the MHC class I epitope-encoding nucleic acid sequences encodes a polypeptide that is between 13 and 25 amino acids in length.

[0071] In some aspects, the LNP comprises a lipid selected from the group consisting of: an ionizable amino lipid, a phosphatidylcholine, cholesterol, a PEG-based coat lipid, or a

combination thereof. In some aspects, the LNP comprises an ionizable amino lipid, a phosphatidylcholine, cholesterol, and a PEG-based coat lipid. In some aspects, the ionizable amino lipids comprise MC3-like (dilinoleylmethyl-4-dimethylaminobutyrate) molecules. In some aspects, the LNP-encapsulated expression system has a diameter of about 100nm.

[0072] In some aspects, the cassette is integrated between the at least one promoter nucleotide sequence and the at least one poly(A) sequence.

[0073] In some aspects, the at least one promoter nucleotide sequence is operably linked to the cassette.

[0074] In some aspects, the one or more vectors comprise one or more +-stranded RNA vectors. In some aspects, the one or more +-stranded RNA vectors comprise a 5' 7-methylguanosine (m7g) cap. In some aspects, the one or more +-stranded RNA vectors are produced by *in vitro* transcription. In some aspects, the one or more vectors are self-replicating within a mammalian cell. In some aspects, the RNA alphavirus backbone comprises at least one nucleotide sequence of an Aura virus, a Fort Morgan virus, a Venezuelan equine encephalitis virus, a Ross River virus, a Semliki Forest virus, a Sindbis virus, or a Mayaro virus. In some aspects, the RNA alphavirus backbone comprises at least one nucleotide sequence of a Venezuelan equine encephalitis virus. In some aspects, the RNA alphavirus backbone comprises at least sequences for nonstructural protein-mediated amplification, a 26S promoter sequence, a poly(A) sequence, a nonstructural protein 1 (nsP1) gene, a nsP2 gene, a nsP3 gene, and a nsP4 gene encoded by the nucleotide sequence of the Aura virus, the Fort Morgan virus, the Venezuelan equine encephalitis virus, the Ross River virus, the Semliki Forest virus, the Sindbis virus, or the Mayaro virus. In some aspects, the RNA alphavirus backbone comprises at least sequences for nonstructural protein-mediated amplification, a 26S promoter sequence, and a poly(A) sequence encoded by the nucleotide sequence of the Aura virus, the Fort Morgan virus, the Venezuelan equine encephalitis virus, the Ross River virus, the Semliki Forest virus, the Sindbis virus, or the Mayaro virus. In some aspects, sequences for nonstructural protein-mediated amplification are selected from the group consisting of: an alphavirus 5' UTR, a 51-nt CSE, a 24-nt CSE, a 26S subgenomic promoter sequence, a 19-nt CSE, an alphavirus 3' UTR, or combinations thereof. In some aspects, the RNA alphavirus backbone does not encode structural virion proteins capsid, E2 and E1. In some aspects, the cassette is inserted in place of structural virion proteins within the nucleotide sequence of the Aura virus, the Fort Morgan virus, the Venezuelan equine encephalitis virus, the Ross River virus, the Semliki Forest virus, the Sindbis virus, or the Mayaro virus. In some aspects, the

Venezuelan equine encephalitis virus comprises the sequence of SEQ ID NO:3 or SEQ ID NO:5. In some aspects, the Venezuelan equine encephalitis virus comprises the sequence of SEQ ID NO:3 or SEQ ID NO:5 further comprising a deletion between base pair 7544 and 11175. In some aspects, the RNA alphavirus backbone comprises the sequence set forth in SEQ ID NO:6 or SEQ ID NO:7. In some aspects, the cassette is inserted at position 7544 to replace the deletion between base pairs 7544 and 11175 as set forth in the sequence of SEQ ID NO:3 or SEQ ID NO:5. In some aspects, the insertion of the cassette provides for transcription of a polycistronic RNA comprising the nsP1-4 genes and the at least one nucleic acid sequence, wherein the nsP1-4 genes and the at least one nucleic acid sequence are in separate open reading frames.

[0075] In some aspects, the at least one promoter nucleotide sequence is the native 26S promoter nucleotide sequence encoded by the RNA alphavirus backbone. In some aspects, the at least one promoter nucleotide sequence is an exogenous RNA promoter. In some aspects, the second promoter nucleotide sequence is a 26S promoter nucleotide sequence. In some aspects, the second promoter nucleotide sequence comprises multiple 26S promoter nucleotide sequences, wherein each 26S promoter nucleotide sequence provides for transcription of one or more of the separate open reading frames.

[0076] In some aspects, the one or more vectors are each at least 300nt in size. In some aspects, the one or more vectors are each at least 1kb in size. In some aspects, the one or more vectors are each 2kb in size. In some aspects, the one or more vectors are each less than 5kb in size.

[0077] In some aspects, the at least one antigen-encoding nucleic acid sequence comprises two or more antigen-encoding nucleic acid sequences. In some aspects, each antigen-encoding nucleic acid sequence is linked directly to one another. In some aspects, each antigen-encoding nucleic acid sequence is linked to a distinct antigen-encoding nucleic acid sequence with a nucleic acid sequence encoding a linker. In some aspects, the linker links two MHC class I epitope-encoding nucleic acid sequences or an MHC class I epitope-encoding nucleic acid sequence to an MHC class II epitope-encoding nucleic acid sequence. In some aspects, the linker is selected from the group consisting of: (1) consecutive glycine residues, at least 2, 3, 4, 5, 6, 7, 8, 9, or 10 residues in length; (2) consecutive alanine residues, at least 2, 3, 4, 5, 6, 7, 8, 9, or 10 residues in length; (3) two arginine residues (RR); (4) alanine, alanine, tyrosine (AAY); (5) a consensus sequence at least 2, 3, 4, 5, 6, 7, 8, 9, or 10 amino acid residues in length that is processed efficiently by a mammalian proteasome; and (6) one or more native

sequences flanking the antigen derived from the cognate protein of origin and that is at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, or 2-20 amino acid residues in length. In some aspects, the linker links two MHC class II epitope-encoding nucleic acid sequences or an MHC class II sequence to an MHC class I epitope-encoding nucleic acid sequence. In some aspects, the linker comprises the sequence GPGPG.

[0078] In some aspects, the antigen-encoding nucleic acid sequences is linked, operably or directly, to a separate or contiguous sequence that enhances the expression, stability, cell trafficking, processing and presentation, and/or immunogenicity of the antigen-encoding nucleic acid sequence. In some aspects, the separate or contiguous sequence comprises at least one of: a ubiquitin sequence, a ubiquitin sequence modified to increase proteasome targeting (e.g., the ubiquitin sequence contains a Gly to Ala substitution at position 76), an immunoglobulin signal sequence (e.g., IgK), a major histocompatibility class I sequence, lysosomal-associated membrane protein (LAMP)-1, human dendritic cell lysosomal-associated membrane protein, and a major histocompatibility class II sequence; optionally wherein the ubiquitin sequence modified to increase proteasome targeting is A76.

[0079] In some aspects, the at least one antigen-encoding nucleic acid sequence comprises at least 2-10, 2, 3, 4, 5, 6, 7, 8, 9, or 10 antigen-encoding nucleic acid sequences, optionally wherein each antigen-encoding nucleic acid sequence encodes a distinct antigen-encoding nucleic acid sequence. In some aspects, the at least one antigen-encoding nucleic acid sequence comprises at least 11-20, 15-20, 11-100, 11-200, 11-300, 11-400, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 or up to 400 antigen-encoding nucleic acid sequences, optionally wherein each antigen-encoding nucleic acid sequence encodes a distinct antigen-encoding nucleic acid sequence. In some aspects, the at least one antigen-encoding nucleic acid sequence comprises at least 11-20, 15-20, 11-100, 11-200, 11-300, 11-400, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 or up to 400 antigen-encoding nucleic acid sequences. In some aspects, the at least one antigen-encoding nucleic acid sequence comprises at least 2-400 antigen-encoding nucleic acid sequences and wherein at least two of the antigen-encoding nucleic acid sequences encode epitope sequences or portions thereof that are presented by MHC class I on a cell surface. In some aspects, at least two of the MHC class I epitopes are presented by MHC class I on the infected cell surface.

[0080] In some aspects, the epitope-encoding nucleic acid sequences comprises at least one MHC class I epitope-encoding nucleic acid sequence, and wherein each antigen-encoding nucleic acid sequence encodes a polypeptide sequence between 8 and 35 amino acids in length,

optionally 9-17, 9-25, 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 or 35 amino acids in length.

[0081] In some aspects, the at least one MHC class II epitope-encoding nucleic acid sequence is present. In some aspects, the at least one MHC class II epitope-encoding nucleic acid sequence is present and comprises at least one MHC class II epitope-encoding nucleic acid sequence that comprises at least one alteration that makes the encoded epitope sequence distinct from the corresponding peptide sequence encoded by a wild-type nucleic acid sequence. In some aspects, the epitope-encoding nucleic acid sequence comprises an MHC class II epitope-encoding nucleic acid sequence and wherein each antigen-encoding nucleic acid sequence encodes a polypeptide sequence that is 12-20, 12, 13, 14, 15, 16, 17, 18, 19, 20, or 20-40 amino acids in length. In some aspects, the epitope-encoding nucleic acid sequences comprises an MHC class II epitope-encoding nucleic acid sequence, wherein the at least one MHC class II epitope-encoding nucleic acid sequence is present, and wherein the at least one MHC class II epitope-encoding nucleic acid sequence comprises at least one universal MHC class II epitope-encoding nucleic acid sequence, optionally wherein the at least one universal sequence comprises at least one of Tetanus toxoid and PADRE.

[0082] In some aspects, the at least one promoter nucleotide sequence or the second promoter nucleotide sequence is inducible. In some aspects, the at least one promoter nucleotide sequence or the second promoter nucleotide sequence is non-inducible. In some aspects, the at least one poly(A) sequence comprises a poly(A) sequence native to the alphavirus. In some aspects, the at least one poly(A) sequence is operably linked to at least one of the at least one nucleic acid sequences. In some aspects, the at least one poly(A) sequence is at least 20, at least 30, at least 40, at least 50, at least 60, at least 70, at least 80, or at least 90 consecutive A nucleotides. In some aspects, the at least one poly(A) sequence is at least 100 consecutive A nucleotides.

[0083] In some aspects, the epitope-encoding nucleic acid sequence comprises a MHC class I epitope-encoding nucleic acid sequence, and wherein the MHC class I epitope-encoding nucleic acid sequence is selected by performing the steps of: (a) obtaining at least one of exome, transcriptome, or whole genome nucleotide sequencing data from the infectious disease organism, wherein the infectious disease organism nucleotide sequencing data is used to obtain data representing peptide sequences of each of a set of epitopes; (b) inputting the peptide sequence of each epitope into a presentation model to generate a set of numerical likelihoods that each of the epitopes is presented by one or more of the MHC alleles on the infected cell

surface, the set of numerical likelihoods having been identified at least based on received mass spectrometry data; and (c) selecting a subset of the set of epitopes based on the set of numerical likelihoods to generate a set of selected epitopes which are used to generate the MHC class I epitope-encoding nucleic acid sequence. In some aspects, each of the MHC class I epitope-encoding nucleic acid sequences is selected by performing the steps of: (a) obtaining at least one of exome, transcriptome, or whole genome infectious disease organism nucleotide sequencing data from the infectious disease organism, wherein the infectious disease organism nucleotide sequencing data is used to obtain data representing peptide sequences of each of a set of epitopes; (b) inputting the peptide sequence of each epitope into a presentation model to generate a set of numerical likelihoods that each of the epitopes is presented by one or more of the MHC alleles on the infected cell surface, the set of numerical likelihoods having been identified at least based on received mass spectrometry data; and (c) selecting a subset of the set of epitopes based on the set of numerical likelihoods to generate a set of selected epitopes which are used to generate the at least 20 MHC class I epitope-encoding nucleic acid sequences. In some aspects, a number of the set of selected epitopes is 2-20. In some aspects, the presentation model represents dependence between: (a) presence of a pair of a particular one of the MHC alleles and a particular amino acid at a particular position of a peptide sequence; and (b) likelihood of presentation on the infected cell surface, by the particular one of the MHC alleles of the pair, of such a peptide sequence comprising the particular amino acid at the particular position. In some aspects, selecting the set of selected epitopes comprises selecting epitopes that have an increased likelihood of being presented on the infected cell surface relative to unselected epitopes based on the presentation model. In some aspects, selecting the set of selected epitopes comprises selecting epitopes that have an increased likelihood of being capable of inducing a infectious disease organism-specific immune response in the subject relative to unselected epitopes based on the presentation model. In some aspects, selecting the set of selected epitopes comprises selecting epitopes that have an increased likelihood of being capable of being presented to naïve T cells by professional antigen presenting cells (APCs) relative to unselected epitopes based on the presentation model, optionally wherein the APC is a dendritic cell (DC). In some aspects, selecting the set of selected epitopes comprises selecting epitopes that have a decreased likelihood of being subject to inhibition via central or peripheral tolerance relative to unselected epitopes based on the presentation model. In some aspects, selecting the set of selected epitopes comprises selecting epitopes that have a decreased likelihood of being capable of inducing an

autoimmune response to normal tissue in the subject relative to unselected epitopes based on the presentation model. In some aspects, exome or transcriptome nucleotide sequencing data is obtained by performing sequencing on the infected tissue. In some aspects, the sequencing is next generation sequencing (NGS) or any massively parallel sequencing approach.

[0084] In some aspects, the ChAdV vector comprises: (a) a ChAdV backbone, wherein the ChAdV backbone comprises: (i) at least one promoter nucleotide sequence, and (ii) at least one polyadenylation (poly(A)) sequence; and (b) a cassette, wherein the cassette comprises: (i) at least one antigen-encoding nucleic acid sequence comprising: a. a nucleic acid sequence encoding an infectious disease organism peptide selected from the group consisting of: a pathogen-derived peptide, a virus-derived peptide, a bacteria-derived peptide, a fungus-derived peptide, and a parasite-derived peptide b. optionally a 5' linker sequence, and c. optionally a 3' linker sequence; and wherein the cassette is operably linked to the at least one promoter nucleotide sequence and the at least one poly(A) sequence.

[0085] In some aspects, the ChAdV vector comprises: (a) a ChAdV backbone, wherein the ChAdV backbone comprises: (i) a modified ChAdV68 sequence comprising at least nucleotides 2 to 36,518 of the sequence set forth in SEQ ID NO:1, wherein the nucleotides 2 to 36,518 lack: (1) nucleotides 577 to 3403 of the sequence shown in SEQ ID NO:1 corresponding to an E1 deletion; (2) nucleotides 27,125 to 31,825 of the sequence shown in SEQ ID NO:1 corresponding to an E3 deletion; and optionally (3) nucleotides 34,916 to 35,642 of the sequence shown in SEQ ID NO:1 corresponding to a partial E4 deletion; (ii) a CMV promoter nucleotide sequence; and (iii) an SV40 polyadenylation (poly(A)) sequence; and (b) a cassette, wherein the cassette comprises: (i) at least one antigen-encoding nucleic acid sequence comprising: a. a nucleic acid sequence encoding an infectious disease organism peptide selected from the group consisting of: a pathogen-derived peptide, a virus-derived peptide, a bacteria-derived peptide, a fungus-derived peptide, and a parasite-derived peptide b. optionally a 5' linker sequence, and c. optionally a 3' linker sequence; and wherein the cassette is inserted within the E1 deletion and the cassette is operably linked to the CMV promoter nucleotide sequence and the SV40 poly(A) sequence.

[0086] In some aspects, the nucleic acid sequence encoding the peptide comprises an epitope-encoding nucleic acid sequence, optionally wherein the nucleic acid sequence encoding the peptide comprises two or more distinct epitope-encoding nucleic acid sequences. In some aspects, the nucleic acid sequence encoding the peptide comprises between 1-10, between 1-20, between 1-30, between 1-40, between 1-50, between 1-100, between 1-200, between 1-300,

between 1-400, or between 1-500 distinct epitope-encoding nucleic acid sequences distinct epitope-encoding nucleic acid sequences, optionally wherein the nucleic acid sequence encoding the peptide comprises between 2-10, between 2-20, between 2-30, between 2-40, between 2-50, between 2-100, between 2-200, between 2-300, between 2-400, or between 2-500 distinct epitope-encoding nucleic acid sequences distinct epitope-encoding nucleic acid sequences. In some aspects, the nucleic acid sequence encoding the peptide comprises a nucleic acid sequence encoding a peptide selected from: an epitope, a full-length protein, a protein subunit, a protein domain, of the protein expressed in the infectious disease organism, optionally wherein the infectious disease organism comprises two or more distinct nucleic acid sequences encoding a peptide selected from: an epitope, a full-length protein, a protein subunit, a protein domain, and combinations thereof of the protein expressed in the infectious disease organism. In some aspects, the encoded peptide or peptides is capable of stimulating an immune response when expressed in a subject, optionally wherein the immune response is a T cell response and/or a B cell response. In some aspects, the infectious disease organism is selected from the group consisting of: Severe acute respiratory syndrome-related coronavirus (SARS), severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), Ebola, HIV, Hepatitis B virus (HBV), influenza, Hepatitis C virus (HCV), Human papillomavirus (HPV), Cytomegalovirus (CMV), Chikungunya virus, Respiratory syncytial virus (RSV), Dengue virus, a orthomyxoviridae family virus, and tuberculosis.

[0087] In some aspects, the epitope-encoding nucleic acid sequence encodes an epitope known or suspected to be presented by MHC class I on a surface of a cell, optionally wherein the surface of the cell is an infected cell surface, and optionally wherein the cell is the subject's cell. In some aspects, the cell is an infected cell selected from the group consisting of: a pathogen infected cell, a virally infected cell, a bacterially infected cell, a fungally infected cell, and a parasitically infected cell. In some aspects, the virally infected cell is selected from the group consisting of: an HIV infected cell, an HPV infected cell, a SARS infected cell, a SARS-CoV-2 infected cell, an Ebola infected cell, a HBV infected cell, an influenza infected cell, a HCV infected cell, a CMV infected cell, a Chikungunya virus infected cell, a RSV infected cell, a Dengue virus infected cell, a orthomyxoviridae family virus infected cell, and a tuberculosis infected cell.

[0088] In some aspects, an ordered sequence of each element of the cassette in the composition for delivery of the ChAdV-based expression system is described in the formula, from 5' to 3', comprising $P_a-(L5_b-N_c-L3_d)_X-(G5_e-U_f)_Y-G3_g$ wherein P comprises the

at least one promoter sequence operably linked to at least one of the at least one antigen-encoding nucleic acid sequences, where $a = 1$, N comprises one of the epitope-encoding nucleic acid sequences, wherein the epitope-encoding nucleic acid sequence comprises an MHC class I epitope-encoding nucleic acid sequence, where $c = 1$, $L5$ comprises the 5' linker sequence, where $b = 0$ or 1 , $L3$ comprises the 3' linker sequence, where $d = 0$ or 1 , $G5$ comprises one of the at least one nucleic acid sequences encoding a GP GPG amino acid linker, where $e = 0$ or 1 , $G3$ comprises one of the at least one nucleic acid sequences encoding a GP GPG amino acid linker, where $g = 0$ or 1 , U comprises one of the at least one MHC class II epitope-encoding nucleic acid sequence, where $f = 1$, $X = 1$ to 400 , where for each X the corresponding N_c is an MHC class I epitope-encoding nucleic acid sequence, and $Y = 0, 1$, or 2 , where for each Y the corresponding U_f is an MHC class II epitope-encoding nucleic acid sequence. In some aspects, for each X the corresponding N_c is a distinct MHC class I epitope-encoding nucleic acid sequence. In some aspects, for each Y the corresponding U_f is a distinct MHC class II epitope-encoding nucleic acid sequence. In some aspects, $b = 1$, $d = 1$, $e = 1$, $g = 1$, $h = 1$, $X = 10$, $Y = 2$, P is a CMV promoter sequence, each N encodes a MHC class I epitope 7-15 amino acids in length, $L5$ is a native 5' linker sequence that encodes a native N-terminal amino acid sequence of the MHC I epitope, and wherein the 5' linker sequence encodes a peptide that is at least 3 amino acids in length, $L3$ is a native 3' linker sequence that encodes a native C-terminal amino acid sequence of the MHC I epitope, and wherein the 3' linker sequence encodes a peptide that is at least 3 amino acids in length, U is each of a PADRE class II sequence and a Tetanus toxoid MHC class II sequence, the ChAdV vector comprises a modified ChAdV68 sequence comprising at least nucleotides 2 to 36,518 of the sequence set forth in SEQ ID NO:1, wherein the nucleotides 2 to 36,518 lack: (1) nucleotides 577 to 3403 of the sequence shown in SEQ ID NO:1 corresponding to an E1 deletion; (2) nucleotides 27,125 to 31,825 of the sequence shown in SEQ ID NO:1 corresponding to an E3 deletion; and optionally (3) nucleotides 34,916 to 35,642 of the sequence shown in SEQ ID NO:1 corresponding to a partial E4 deletion, and the antigen cassette is inserted within the E1 deletion, and each of the MHC class I antigen-encoding nucleic acid sequences encodes a polypeptide that is 25 amino acids in length.

[0089] In some aspects, the cassette is integrated between the at least one promoter nucleotide sequence and the at least one poly(A) sequence. In some aspects, the at least one promoter nucleotide sequence is operably linked to the cassette.

[0090] In some aspects, the ChAdV backbone comprises a ChAdV68 vector backbone. In some aspects, the ChAdV68 vector backbone comprises the sequence set forth in SEQ ID

NO:1. In some aspects, the ChAdV68 vector backbone comprises a functional deletion in at least one gene selected from the group consisting of an adenovirus E1A, E1B, E2A, E2B, E3, L1, L2, L3, L4, and L5 gene with reference to a ChAdV68 genome or with reference to the sequence shown in SEQ ID NO:1, optionally wherein the adenoviral backbone or modified ChAdV68 sequence is fully deleted or functionally deleted in: (1) E1A and E1B; or (2) E1A, E1B, and E3 with reference to the adenovirus genome or with reference to the sequence shown in SEQ ID NO:1, optionally wherein the E1 gene is functionally deleted through an E1 deletion of at least nucleotides 577 to 3403 with reference to the sequence shown in SEQ ID NO:1 and optionally wherein the E3 gene is functionally deleted through an E3 deletion of at least nucleotides 27,125 to 31,825 with reference to the sequence shown in SEQ ID NO:1. In some aspects, the ChAdV68 vector backbone comprises one or more genes or regulatory sequences with reference to a ChAdV68 genome or with reference to the sequence shown in SEQ ID NO:1, optionally wherein the one or more genes or regulatory sequences are selected from the group consisting of the chimpanzee adenovirus inverted terminal repeat (ITR), E1A, E1B, E2A, E2B, E3, E4, L1, L2, L3, L4, and L5 genes.

[0091] In some aspects, the ChAdV68 vector backbone comprises a partially deleted E4 gene. In some aspects, the the partially deleted E4 gene comprises: A. the E4 gene sequence shown in SEQ ID NO:1 and that lacks at least nucleotides 34,916 to 35,642 of the sequence shown in SEQ ID NO:1, B. the E4 gene sequence shown in SEQ ID NO:1 and that lacks at least nucleotides 34,916 to 34,942, nucleotides 34,952 to 35,305 of the sequence shown in SEQ ID NO:1, nucleotides 35,302 to 35,642 of the sequence shown in SEQ ID NO:1, and wherein the vector comprises at least nucleotides 2 to 36,518 of the sequence shown in SEQ ID NO:1, C. the E4 gene sequence shown in SEQ ID NO:1 and that lacks at least nucleotides 34,980 to 36,516 of the sequence shown in SEQ ID NO:1, and wherein the vector comprises at least nucleotides 2 to 36,518 of the sequence shown in SEQ ID NO:1, D. the E4 gene sequence shown in SEQ ID NO:1 and that lacks at least nucleotides 34,979 to 35,642 of the sequence shown in SEQ ID NO:1, and wherein the vector comprises at least nucleotides 2 to 36,518 of the sequence shown in SEQ ID NO:1, E. an E4 deletion of at least a partial deletion of E4Orf2, a fully deleted E4Orf3, and at least a partial deletion of E4Orf4, F. an E4 deletion of at least a partial deletion of E4Orf2, at least a partial deletion of E4Orf3, and at least a partial deletion of E4Orf4, G. an E4 deletion of at least a partial deletion of E4Orf1, a fully deleted E4Orf2, and at least a partial deletion of E4Orf3, or H. an E4 deletion of at least a partial deletion of E4Orf2 and at least a partial deletion of E4Orf3. In some aspects, the ChAdV68 vector backbone

comprises at least nucleotides 2 to 36,518 of the sequence set forth in SEQ ID NO:1, wherein the nucleotides 2 to 36,518 lack: (1) nucleotides 577 to 3403 of the sequence shown in SEQ ID NO:1 corresponding to an E1 deletion; (2) nucleotides 27,125 to 31,825 of the sequence shown in SEQ ID NO:1 corresponding to an E3 deletion; and (3) nucleotides 34,916 to 35,642 of the sequence shown in SEQ ID NO:1 corresponding to a partial E4 deletion; optionally wherein the antigen cassette is inserted within the E1 deletion. In some aspects, the ChAdV68 vector backbone comprises the sequence set forth in SEQ ID NO: 29369, optionally wherein the antigen cassette is inserted within the E1 deletion. In some aspects, the ChAdV68 vector backbone comprises at least nucleotides 2 to 36,518 of the sequence set forth in SEQ ID NO:1, wherein the nucleotides 2 to 36,518 lack: A. nucleotides 577 to 3403 of the sequence shown in SEQ ID NO:1 corresponding to an E1 deletion; B. nucleotides 27,125 to 31,825 of the sequence shown in SEQ ID NO:1 corresponding to an E3 deletion; C. nucleotides 34,916 to 35,642 of the sequence shown in SEQ ID NO:1 corresponding to a partial E4 deletion; D. nucleotides 456 to 3014 of the sequence shown in SEQ ID NO:1; E. nucleotides 27,816 to 31,333 of the sequence shown in SEQ ID NO:1; F. nucleotides 3957 to 10346 of the sequence shown in SEQ ID NO:1; G. nucleotides 21787 to 23370 of the sequence shown in SEQ ID NO:1; H. nucleotides 33486 to 36193 of the sequence shown in SEQ ID NO:1; or combinations thereof. In some aspects, the ChAdV68 vector backbone comprises at least nucleotides 2 to 36,518 of the sequence set forth in SEQ ID NO:1, wherein the nucleotides 2 to 36,518 lack: (1) nucleotides 577 to 3403 of the sequence shown in SEQ ID NO:1 corresponding to an E1 deletion and (2) nucleotides 27,125 to 31,825 of the sequence shown in SEQ ID NO:1 corresponding to an E3 deletion.

[0092] In some aspects, the ChAdV68 vector backbone comprises the sequence set forth in SEQ ID NO:1, except that the sequence is fully deleted or functionally deleted in at least one gene selected from the group consisting of the chimpanzee adenovirus E1A, E1B, E2A, E2B, E3, E4, L1, L2, L3, L4, and L5 genes of the sequence set forth in SEQ ID NO: 1, optionally wherein the sequence is fully deleted or functionally deleted in: (1) E1A and E1B; (2) E1A, E1B, and E3; or (3) E1A, E1B, E3, and E4 of the sequence set forth in SEQ ID NO: 1. In some aspects, the ChAdV68 vector backbone comprises a gene or regulatory sequence obtained from the sequence of SEQ ID NO: 1, optionally wherein the gene is selected from the group consisting of the chimpanzee adenovirus inverted terminal repeat (ITR), E1A, E1B, E2A, E2B, E3, E4, L1, L2, L3, L4, and L5 genes of the sequence set forth in SEQ ID NO: 1. In some aspects, the ChAdV68 vector backbone comprises one or more deletions between base pair

number 577 and 3403 or between base pair 456 and 3014, and optionally wherein the vector further comprises one or more deletions between base pair 27,125 and 31,825 or between base pair 27,816 and 31,333 of the sequence set forth in SEQ ID NO:1. In some aspects, the ChAdV68 vector backbone comprises one or more deletions between base pair number 3957 and 10346, base pair number 21787 and 23370, and base pair number 33486 and 36193 of the sequence set forth in SEQ ID NO:1.

[0093] In some aspects, the cassette is inserted in the ChAdV backbone at the E1 region, E3 region, and/or any deleted AdV region that allows incorporation of the cassette. In some aspects, the ChAdV backbone is generated from one of a first generation, a second generation, or a helper-dependent adenoviral vector.

[0094] In some aspects, the at least one promoter nucleotide sequence is selected from the group consisting of: a CMV, a SV40, an EF-1, a RSV, a PGK, a HSA, a MCK, and a EBV promoter sequence. In some aspects, the at least one promoter nucleotide sequence is a CMV promoter sequence.

[0095] In some aspects, at least one of the epitope-encoding nucleic acid sequences encodes an epitope that, when expressed and translated, is capable of being presented by MHC class I on a cell of the subject. In some aspects, at least one of the epitope-encoding nucleic acid sequences encodes an epitope that, when expressed and translated, is capable of being presented by MHC class II on a cell of the subject.

[0096] In some aspects, the at least one antigen-encoding nucleic acid sequence comprises two or more antigen-encoding nucleic acid sequences. In some aspects, each antigen-encoding nucleic acid sequence is linked directly to one another.

[0097] In some aspects, each antigen-encoding nucleic acid sequence is linked to a distinct antigen-encoding nucleic acid sequence with a nucleic acid sequence encoding a linker. In some aspects, the linker links two MHC class I epitope-encoding nucleic acid sequences or an MHC class I epitope-encoding nucleic acid sequence to an MHC class II epitope-encoding nucleic acid sequence. In some aspects, the linker is selected from the group consisting of: (1) consecutive glycine residues, at least 2, 3, 4, 5, 6, 7, 8, 9, or 10 residues in length; (2) consecutive alanine residues, at least 2, 3, 4, 5, 6, 7, 8, 9, or 10 residues in length; (3) two arginine residues (RR); (4) alanine, alanine, tyrosine (AAY); (5) a consensus sequence at least 2, 3, 4, 5, 6, 7, 8, 9, or 10 amino acid residues in length that is processed efficiently by a mammalian proteasome; and (6) one or more native sequences flanking the antigen derived from the cognate protein of origin and that is at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15,

16, 17, 18, 19, 20, or 2-20 amino acid residues in length. In some aspects, the linker links two MHC class II epitope-encoding nucleic acid sequences or an MHC class II sequence to an MHC class I epitope-encoding nucleic acid sequence. In some aspects, the linker comprises the sequence GP GPG.

[0098] In some aspects, the antigen-encoding nucleic acid sequences is linked, operably or directly, to a separate or contiguous sequence that enhances the expression, stability, cell trafficking, processing and presentation, and/or immunogenicity of the antigen-encoding nucleic acid sequence. In some aspects, the separate or contiguous sequence comprises at least one of: a ubiquitin sequence, a ubiquitin sequence modified to increase proteasome targeting (e.g., the ubiquitin sequence contains a Gly to Ala substitution at position 76), an immunoglobulin signal sequence (e.g., IgK), a major histocompatibility class I sequence, lysosomal-associated membrane protein (LAMP)-1, human dendritic cell lysosomal-associated membrane protein, and a major histocompatibility class II sequence; optionally wherein the ubiquitin sequence modified to increase proteasome targeting is A76.

[0099] In some aspects, the epitope-encoding nucleic acid sequence comprises at least one alteration that makes the encoded epitope have increased binding affinity to its corresponding MHC allele relative to the translated, corresponding wild-type nucleic acid sequence. In some aspects, the epitope-encoding nucleic acid sequence comprises at least one alteration that makes the encoded epitope have increased binding stability to its corresponding MHC allele relative to the translated, corresponding wild-type nucleic acid sequence. In some aspects, the epitope-encoding nucleic acid sequence comprises at least one alteration that makes the encoded epitope have an increased likelihood of presentation on its corresponding MHC allele relative to the translated, corresponding wild-type nucleic acid sequence. In some aspects, the at least one alteration comprises a point mutation, a frameshift mutation, a non-frameshift mutation, a deletion mutation, an insertion mutation, a splice variant, a genomic rearrangement, or a proteasome-generated spliced antigen.

[00100] In some aspects, the epitope-encoding nucleic acid sequence encodes an epitope known or suspected to be expressed in the subject known or suspected to have an infection. In some aspects, the infection is selected from the group consisting of: a pathogen infection, a viral infection, a bacterial infection, an fungal infection, and a parasitic infection. In some aspects, the viral infection is selected from the group consisting of: an HIV infection, an HPV infection, a SARS infection, a SARS-CoV-2 infection, an Ebola infection, a HBV infection, an influenza infection, a HCV infection, a CMV infection, a Chikungunya virus infection, a RSV

infection, a Dengue virus infection, a orthomyxoviridae family virus infection, and a tuberculosis infection.

[00101] In some aspects, the at least one antigen-encoding nucleic acid sequence comprises at least 2-10, 2, 3, 4, 5, 6, 7, 8, 9, or 10 antigen-encoding nucleic acid sequences, optionally wherein each antigen-encoding nucleic acid sequence encodes a distinct antigen-encoding nucleic acid sequence. In some aspects, the at least one antigen-encoding nucleic acid sequence comprises at least 11-20, 15-20, 11-100, 11-200, 11-300, 11-400, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 or up to 400 antigen-encoding nucleic acid sequences, optionally wherein each antigen-encoding nucleic acid sequence encodes a distinct antigen-encoding nucleic acid sequence. In some aspects, the at least one antigen-encoding nucleic acid sequence comprises at least 11-20, 15-20, 11-100, 11-200, 11-300, 11-400, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 or up to 400 antigen-encoding nucleic acid sequences. In some aspects, the at least one antigen-encoding nucleic acid sequence comprises at least 2-400 antigen-encoding nucleic acid sequences and wherein at least two of the antigen-encoding nucleic acid sequences encode epitope sequences or portions thereof that are presented by MHC class I on a cell surface. In some aspects, at least two of the MHC class I epitopes are presented by MHC class I on the infected cell surface.

[00102] In some aspects, the epitope-encoding nucleic acid sequences comprises at least one MHC class I epitope-encoding nucleic acid sequence, and wherein each antigen-encoding nucleic acid sequence encodes a polypeptide sequence between 8 and 35 amino acids in length, optionally 9-17, 9-25, 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 or 35 amino acids in length.

[00103] In some aspects, the at least one MHC class II epitope-encoding nucleic acid sequence is present. In some aspects, the at least one MHC class II epitope-encoding nucleic acid sequence is present and comprises at least one MHC class II epitope-encoding nucleic acid sequence that comprises at least one alteration that makes the encoded epitope sequence distinct from the corresponding peptide sequence encoded by a wild-type nucleic acid sequence. In some aspects, the epitope-encoding nucleic acid sequence comprises an MHC class II epitope-encoding nucleic acid sequence and wherein each antigen-encoding nucleic acid sequence encodes a polypeptide sequence that is 12-20, 12, 13, 14, 15, 16, 17, 18, 19, 20, or 20-40 amino acids in length. In some aspects, the epitope-encoding nucleic acid sequences comprises an MHC class II epitope-encoding nucleic acid sequence, wherein the at least one MHC class II epitope-encoding nucleic acid sequence is present, and wherein the at least one MHC class II epitope-encoding nucleic acid sequence comprises at least one universal MHC

class II epitope-encoding nucleic acid sequence, optionally wherein the at least one universal sequence comprises at least one of Tetanus toxoid and PADRE.

[00104] In some aspects, the at least one promoter nucleotide sequence is inducible. In some aspects, wherein the at least one promoter nucleotide sequence is non-inducible. In some aspects, the at least one poly(A) sequence comprises a Bovine Growth Hormone (BGH) SV40 polyA sequence. In some aspects, the at least one poly(A) sequence is at least 20, at least 30, at least 40, at least 50, at least 60, at least 70, at least 80, or at least 90 consecutive A nucleotides. In some aspects, the at least one poly(A) sequence is at least 100 consecutive A nucleotides.

[00105] In some aspects, the cassette further comprises at least one of: an intron sequence, a woodchuck hepatitis virus posttranscriptional regulatory element (WPRE) sequence, an internal ribosome entry sequence (IRES) sequence, a nucleotide sequence encoding a 2A self cleaving peptide sequence, a nucleotide sequence encoding a Furin cleavage site, or a sequence in the 5' or 3' non-coding region known to enhance the nuclear export, stability, or translation efficiency of mRNA that is operably linked to at least one of the at least one antigen-encoding nucleic acid sequences. In some aspects, the cassette further comprises a reporter gene, including but not limited to, green fluorescent protein (GFP), a GFP variant, secreted alkaline phosphatase, luciferase, a luciferase variant, or a detectable peptide or epitope. In some aspects, the detectable peptide or epitope is selected from the group consisting of an HA tag, a Flag tag, a His-tag, or a V5 tag.

[00106] In some aspects, the one or more vectors further comprises one or more nucleic acid sequences encoding at least one immune modulator. In some aspects, the immune modulator is an anti-CTLA4 antibody or an antigen-binding fragment thereof, an anti-PD-1 antibody or an antigen-binding fragment thereof, an anti-PD-L1 antibody or an antigen-binding fragment thereof, an anti-4-1BB antibody or an antigen-binding fragment thereof, or an anti-OX-40 antibody or an antigen-binding fragment thereof. In some aspects, the antibody or antigen-binding fragment thereof is a Fab fragment, a Fab' fragment, a single chain Fv (scFv), a single domain antibody (sdAb) either as single specific or multiple specificities linked together (e.g., camelid antibody domains), or full-length single-chain antibody (e.g., full-length IgG with heavy and light chains linked by a flexible linker). In some aspects, the heavy and light chain sequences of the antibody are a contiguous sequence separated by either a self-cleaving sequence such as 2A or IRES; or the heavy and light chain sequences of the antibody are linked by a flexible linker such as consecutive glycine residues. In some aspects, the immune

modulator is a cytokine. In some aspects, the cytokine is at least one of IL-2, IL-7, IL-12, IL-15, or IL-21 or variants thereof of each.

[00107] In some aspects, the epitope-encoding nucleic acid sequence comprises a MHC class I epitope-encoding nucleic acid sequence, and wherein the MHC class I epitope-encoding nucleic acid sequence is selected by performing the steps of: (a) obtaining at least one of exome, transcriptome, or whole genome infectious disease organism nucleotide sequencing data from the infectious disease organism, wherein the infectious disease organism nucleotide sequencing data is used to obtain data representing peptide sequences of each of a set of epitopes; (b) inputting the peptide sequence of each epitope into a presentation model to generate a set of numerical likelihoods that each of the epitopes is presented by one or more of the MHC alleles on the infected cell surface, the set of numerical likelihoods having been identified at least based on received mass spectrometry data; and (c) selecting a subset of the set of epitopes based on the set of numerical likelihoods to generate a set of selected epitopes which are used to generate the MHC class I epitope-encoding nucleic acid sequence. In some aspects, each of the MHC class I epitope-encoding nucleic acid sequences is selected by performing the steps of: (a) obtaining at least one of exome, transcriptome, or whole genome infectious disease organism nucleotide sequencing data from the infectious disease organism, wherein the infectious disease organism nucleotide sequencing data is used to obtain data representing peptide sequences of each of a set of epitopes; (b) inputting the peptide sequence of each epitope into a presentation model to generate a set of numerical likelihoods that each of the epitopes is presented by one or more of the MHC alleles on the infected cell surface, the set of numerical likelihoods having been identified at least based on received mass spectrometry data; and (c) selecting a subset of the set of epitopes based on the set of numerical likelihoods to generate a set of selected epitopes which are used to generate the at least 20 MHC class I epitope-encoding nucleic acid sequences. In some aspects, a number of the set of selected epitopes is 2-20. In some aspects, the presentation model represents dependence between: (a) presence of a pair of a particular one of the MHC alleles and a particular amino acid at a particular position of a peptide sequence; and (b) likelihood of presentation on the infected cell surface, by the particular one of the MHC alleles of the pair, of such a peptide sequence comprising the particular amino acid at the particular position. In some aspects, selecting the set of selected epitopes comprises selecting epitopes that have an increased likelihood of being presented on the infected cell surface relative to unselected epitopes based on the presentation model. In some aspects, selecting the set of selected epitopes comprises selecting epitopes that

have an increased likelihood of being capable of inducing a infectious disease organism-specific immune response in the subject relative to unselected epitopes based on the presentation model. In some aspects, selecting the set of selected epitopes comprises selecting epitopes that have an increased likelihood of being capable of being presented to naïve T cells by professional antigen presenting cells (APCs) relative to unselected epitopes based on the presentation model, optionally wherein the APC is a dendritic cell (DC). In some aspects, selecting the set of selected epitopes comprises selecting epitopes that have a decreased likelihood of being subject to inhibition via central or peripheral tolerance relative to unselected epitopes based on the presentation model. In some aspects, selecting the set of selected epitopes comprises selecting epitopes that have a decreased likelihood of being capable of inducing an autoimmune response to normal tissue in the subject relative to unselected epitopes based on the presentation model. In some aspects, exome or transcriptome nucleotide sequencing data is obtained by performing sequencing on the infected tissue. In some aspects, the sequencing is next generation sequencing (NGS) or any massively parallel sequencing approach.

[00108] In some aspects, the cassette comprises junctional epitope sequences formed by adjacent sequences in the cassette. In some aspects, at least one or each junctional epitope sequence has an affinity of greater than 500 nM for MHC. In some aspects, each junctional epitope sequence is non-self.

[00109] In some aspects, the cassette does not encode a non-therapeutic MHC class I or class II epitope nucleic acid sequence comprising a translated, wild-type nucleic acid sequence, wherein the non-therapeutic epitope is predicted to be displayed on an MHC allele of the subject. In some aspects, the non-therapeutic predicted MHC class I or class II epitope sequence is a junctional epitope sequence formed by adjacent sequences in the cassette. In some aspects, the prediction is based on presentation likelihoods generated by inputting sequences of the non-therapeutic epitopes into a presentation model. In some aspects, an order of the antigen-encoding nucleic acid sequences in the cassette is determined by a series of steps comprising: (a) generating a set of candidate cassette sequences corresponding to different orders of the antigen-encoding nucleic acid sequences; (b) determining, for each candidate cassette sequence, a presentation score based on presentation of non-therapeutic epitopes in the candidate cassette sequence; and (c) selecting a candidate cassette sequence associated with a presentation score below a predetermined threshold as the cassette sequence for a vaccine.

[00110] In some aspects, the composition for delivery of the ChAdV-based expression system is formulated in a pharmaceutical composition comprising a pharmaceutically acceptable carrier.

[00111] In some aspects, one or more of the epitope-encoding nucleic acid sequences are derived from an infection in or an infected cell of a subject. In some aspects, each of the epitope-encoding nucleic acid sequences are derived from an infection in or an infected cell of a subject. In some aspects, one or more of the epitope-encoding nucleic acid sequences are not derived from an infection in or an infected cell of a subject. In some aspects, each of the epitope-encoding nucleic acid sequences are not derived from an infection in or an infected cell of a subject.

[00112] Also provided for herein is a method for stimulating an immune response in a subject, the method comprising administering to the subject a composition for delivery of a self-replicating alphavirus-based expression system: wherein the composition for delivery of the self-replicating alphavirus-based expression system comprises: (A) the self-replicating alphavirus-based expression system, wherein the self-replicating alphavirus-based expression system comprises one or more vectors, wherein the one or more vectors comprises: (a) an RNA alphavirus backbone, wherein the RNA alphavirus backbone comprises: (i) at least one promoter nucleotide sequence, and (ii) at least one polyadenylation (poly(A)) sequence; and (b) a cassette, wherein the cassette comprises: (i) at least one antigen-encoding nucleic acid sequence comprising: a. a nucleic acid sequence encoding an infectious disease organism peptide selected from the group consisting of: a pathogen-derived peptide, a virus-derived peptide, a bacteria-derived peptide, a fungus-derived peptide, and a parasite-derived peptide, wherein the infectious disease organism is selected from the group consisting of: Severe acute respiratory syndrome-related coronavirus (SARS), severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), Ebola, HIV, Hepatitis B virus (HBV), influenza, Hepatitis C virus (HCV), Human papillomavirus (HPV), Cytomegalovirus (CMV), Chikungunya virus, Respiratory syncytial virus (RSV), Dengue virus, a orthomyxoviridae family virus, and tuberculosis; b. optionally a 5' linker sequence, and c. optionally a 3' linker sequence; (ii) optionally, a second promoter nucleotide sequence operably linked to the at least one antigen-encoding nucleic acid sequence; and (iii) optionally, at least one second poly(A) sequence, wherein the second poly(A) sequence is a native poly(A) sequence or an exogenous poly(A) sequence to the alphavirus, and (B) a lipid-nanoparticle (LNP), wherein the LNP encapsulates the self-replicating alphavirus-based expression system.

[00113] Also provided for herein is a method for stimulating an immune response in a subject, the method comprising administering to the subject a composition for delivery of a chimpanzee adenovirus (ChAdV)-based expression system, wherein the composition for delivery of the ChAdV-based expression system comprises: the ChAdV-based expression system, wherein the ChAdV-based expression system comprises a viral particle comprising a ChAdV vector, wherein the ChAdV vector comprises: (a) a ChAdV backbone, wherein the ChAdV backbone comprises: (i) at least one promoter nucleotide sequence, and (ii) at least one polyadenylation (poly(A)) sequence; and (b) a cassette, wherein the cassette comprises: (i) at least one antigen-encoding nucleic acid sequence comprising: a. a nucleic acid sequence encoding an infectious disease organism peptide selected from the group consisting of: a pathogen-derived peptide, a virus-derived peptide, a bacteria-derived peptide, a fungus-derived peptide, and/or a parasite-derived peptide, wherein the infectious disease organism is selected from the group consisting of: Severe acute respiratory syndrome-related coronavirus (SARS), severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), Ebola, HIV, Hepatitis B virus (HBV), influenza, Hepatitis C virus (HCV), Human papillomavirus (HPV), Cytomegalovirus (CMV), Chikungunya virus, Respiratory syncytial virus (RSV), Dengue virus, a orthomyxoviridae family virus, and tuberculosis; b. optionally a 5' linker sequence, and c. optionally a 3' linker sequence; and wherein the cassette is operably linked to the at least one promoter nucleotide sequence and the at least one poly(A) sequence.

[00114] In some aspects, the self-replicating alphavirus-based expression system or the ChAdV-based expression system comprise any of the features of any one of the above claims.

[00115] The method of any of the above claims, wherein the cassette of the composition for delivery of the ChAdV-based expression system is identical to the cassette of the composition for delivery of the self-replicating alphavirus-based expression system.

BRIEF DESCRIPTION OF THE SEVERAL VIEWS OF THE DRAWINGS

[00116] These and other features, aspects, and advantages of the present invention will become better understood with regard to the following description, and accompanying drawings, where:

[00117] **FIG. 1** illustrates antigen-specific cellular immune responses measured using ELISpot. Antigen-specific IFN-gamma production to six different mamu A01 restricted epitopes was measured in PBMCs for the VEE-MAG25mer srRNA-LNP1(30 µg) (**FIG. 1A**),

VEE-MAG25mer srRNA-LNP1(100 µg) (**FIG. 1B**), or VEE-MAG25mer srRNA-LNP2(100 µg) (**FIG. 1C**) homologous prime/boost or the ChAdV68.5WTnt.MAG25mer /VEE-MAG25mer srRNA heterologous prime/boost group (**FIG. 1D**) using ELISpot 1, 2, 3, 4, 5, 6, 8, 9, or 10 weeks after the first boost immunization (6 rhesus macaques per group). Results are presented as mean spot forming cells (SFC) per 10^6 PBMCs for each epitope in a stacked bar graph format. Values for each animal were normalized to the levels at pre-bleed (week 0).

[00118] **FIG. 2** shows antigen-specific cellular immune response measured using ELISpot. Antigen-specific IFN-gamma production to six different mamu A01 restricted epitopes was measured in PBMCs after immunization with the ChAdV68.5WTnt.MAG25mer /VEE-MAG25mer srRNA heterologous prime/boost regimen using ELISpot prior to immunization and 4, 5, 6, 7, 8, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23 or 24 weeks after the initial immunization. Results are presented as mean spot forming cells (SFC) per 10^6 PBMCs for each epitope (6 rhesus macaques per group) in a stacked bar graph format.

[00119] **FIG. 3** shows antigen-specific cellular immune response measured using ELISpot. Antigen-specific IFN-gamma production to six different mamu A01 restricted epitopes was measured in PBMCs after immunization with the VEE-MAG25mer srRNA LNP2 homologous prime/boost regimen using ELISpot prior to immunization and 4, 5, 6, 7, 8, 10, 11, 12, 13, 14, or 15 weeks after the initial immunization. Results are presented as mean spot forming cells (SFC) per 10^6 PBMCs for each epitope (6 rhesus macaques per group) in a stacked bar graph format.

[00120] **FIG. 4** shows antigen-specific cellular immune response measured using ELISpot. Antigen-specific IFN-gamma production to six different mamu A01 restricted epitopes was measured in PBMCs after immunization with the VEE-MAG25mer srRNA LNP1 homologous prime/boost regimen using ELISpot prior to immunization and 4, 5, 6, 7, 8, 10, 11, 12, 13, 14, or 15 weeks after the initial immunization. Results are presented as mean spot forming cells (SFC) per 10^6 PBMCs for each epitope (6 rhesus macaques per group) in a stacked bar graph format.

[00121] **FIG. 5** illustrates the vaccination strategy used to evaluate immunogenicity of the antigen-cassette containing vectors in rhesus macaques. Triangles indicate chAd68 vaccination (1×10^{12} vp/animal) at weeks 0 & 32. Circles represent alphavirus vaccination at weeks 0, 4, 12, 20, 28 & 32. Squares represent administration of an anti-CTLA4 antibody.

[00122] **FIG. 6** shows a time course of CD8⁺ anti-epitope responses in Rhesus Macaques dosed with chAd-MAG alone (Group 4). Mean SFC/ 1×10^6 splenocytes is shown.

[00123] FIG. 7 shows a time course of CD8⁺ anti-epitope responses in Rhesus Macaques dosed with chAd-MAG plus anti-CTLA4 antibody (Ipilimumab) delivered IV. (Group 5). Mean SFC/1e6 splenocytes is shown.

[00124] FIG. 8 shows a time course of CD8⁺ anti-epitope responses in Rhesus Macaques dosed with chAd-MAG plus anti-CTLA4 antibody (Ipilimumab) delivered SC (Group 6). Mean SFC/1e6 splenocytes is shown.

[00125] FIG. 9 shows antigen-specific memory responses generated by ChAdV68/samRNA vaccine protocol measured by ELISpot. Results are presented as individual dot plots, with each dot representing a single animal. Pre-immunization baseline (left panel) and memory response at 18 months post-prime (right panel) are shown.

[00126] FIG. 10 shows memory cell phenotyping of antigen-specific CD8⁺ T-cells by flow cytometry using combinatorial tetramer staining and CD45RA/CCR7 co-staining.

[00127] FIG. 11 shows the distribution of memory cell types within the sum of the four Mamu-A*01 tetramer⁺ CD8⁺ T-cell populations at study month 18. Memory cells were characterized as follows: CD45RA⁺CCR7⁺=naïve, CD45RA⁺CCR7⁻=effector (Teff), CD45RA⁻CCR7⁺=central memory (Tcm), CD45RA⁻CCR7⁻=effector memory (Tem).

[00128] FIG. 12A shows HAI responses against influenza A H1N1 strain CA/7/09 post dosing with 5e10 vp/animals. HAI 5 is the limit of detection. HAI 40 is a therapeutic threshold.

[00129] FIG. 12B shows HAI responses against influenza A H1N1 strain CA/7/09 post dosing with (B) 1e9 vp/animal. HAI 5 is the limit of detection. HAI 40 is a therapeutic threshold.

[00130] FIG. 13A shows HAI responses against influenza A H7N9 strain Anhui/1/13 post dosing with 5e10 vp/animal. HAI 5 is the limit of detection. HAI 40 is a therapeutic threshold. Note the vaccine is against Anhui/1/13 but the strain used in the HAI assay is the closely related A/Shanghai/2/13 HA on a PR8 influenza strain.

[00131] FIG. 13B shows HAI responses against influenza A H7N9 strain Anhui/1/13 post dosing with 1e9 vp/animal. HAI 5 is the limit of detection. HAI 40 is a therapeutic threshold. Note the vaccine is against Anhui/1/13 but the strain used in the HAI assay is the closely related A/Shanghai/2/13 HA on a PR8 influenza strain.

[00132] FIG. 14A shows anti-HA T-cell response for each mouse (n = 6/group) 2-weeks post ChAd prime. Splenocytes were isolated ex vivo overnight IFN γ ELISpot performed using six overlapping peptide pools spanning the HA antigen. SFU/1e6 splenocytes, response to six pools are stacked.

[00133] FIG. 14B shows anti-HA T-cell response for each mouse ($n = 6/\text{group}$) 2-weeks post SAM prime. Splenocytes were isolated ex vivo overnight IFN γ ELISpot performed using six overlapping peptide pools spanning the HA antigen. SFU/1e6 splenocytes, response to six pools are stacked.

[00134] FIG. 14C shows anti-HA T-cell response for each mouse ($n = 6/\text{group}$) 2-weeks post boost with SAM-HA (10-weeks post ChAd prime) (C). Splenocytes were isolated ex vivo overnight IFN γ ELISpot performed using six overlapping peptide pools spanning the HA antigen. SFU/1e6 splenocytes, response to six pools are stacked.

[00135] FIG. 15A shows T-cell responses post chAd prime/ SAM boost (2-weeks post boost, 10 weeks post prime). Percentage of CD8 $^{+}$ expressing IFN γ , TNF α and IL2 as measured by ICS post 5h ex vivo stimulation (sum of two overlapping peptide pools). Box and whiskers and median, IQR and range.

[00136] FIG. 15B shows T-cell responses post chAd prime/ SAM boost (2-weeks post boost, 10 weeks post prime). Percentage of CD4 $^{+}$ cells expressing IFN γ , TNF α and IL2 as measured by ICS post 5h ex vivo stimulation (sum of two overlapping peptide pools). Box and whiskers and median, IQR and range.

I. Definitions

[00137] In general, terms used in the claims and the specification are intended to be construed as having the plain meaning understood by a person of ordinary skill in the art. Certain terms are defined below to provide additional clarity. In case of conflict between the plain meaning and the provided definitions, the provided definitions are to be used.

[00138] As used herein the term “antigen” is a substance that induces an immune response. An antigen can be a neoantigen. An antigen can be a “shared antigen” that is an antigen found among a specific population, *e.g.*, a specific population of patients infected with or at risk of infection for an infectious disease.

[00139] an antigenAs used herein the term “antigen-based vaccine” is a vaccine composition based on one or more antigens, *e.g.*, a plurality of antigens. The vaccines can be nucleotide-based (*e.g.*, virally based, RNA based, or DNA based), protein-based (*e.g.*, peptide based), or a combination thereof.

[00140] As used herein the term “candidate antigen” is a mutation or other aberration giving rise to a sequence that may represent an antigen.

[00141] As used herein the term “coding region” is the portion(s) of a gene that encode protein.

[00142] As used herein the term “coding mutation” is a mutation occurring in a coding region.

[00143] As used herein the term “ORF” means open reading frame.

[00144] As used herein the term “missense mutation” is a mutation causing a substitution from one amino acid to another.

[00145] As used herein the term “nonsense mutation” is a mutation causing a substitution from an amino acid to a stop codon or causing removal of a canonical start codon.

[00146] As used herein the term “frameshift mutation” is a mutation causing a change in the frame of the protein.

[00147] As used herein the term “indel” is an insertion or deletion of one or more nucleic acids.

[00148] As used herein, the term percent "identity," in the context of two or more nucleic acid or polypeptide sequences, refer to two or more sequences or subsequences that have a specified percentage of nucleotides or amino acid residues that are the same, when compared and aligned for maximum correspondence, as measured using one of the sequence comparison algorithms described below (e.g., BLASTP and BLASTN or other algorithms available to persons of skill) or by visual inspection. Depending on the application, the percent "identity" can exist over a region of the sequence being compared, e.g., over a functional domain, or, alternatively, exist over the full length of the two sequences to be compared.

[00149] For sequence comparison, typically one sequence acts as a reference sequence to which test sequences are compared. When using a sequence comparison algorithm, test and reference sequences are input into a computer, subsequence coordinates are designated, if necessary, and sequence algorithm program parameters are designated. The sequence comparison algorithm then calculates the percent sequence identity for the test sequence(s) relative to the reference sequence, based on the designated program parameters. Alternatively, sequence similarity or dissimilarity can be established by the combined presence or absence of particular nucleotides, or, for translated sequences, amino acids at selected sequence positions (e.g., sequence motifs).

[00150] Optimal alignment of sequences for comparison can be conducted, e.g., by the local homology algorithm of Smith & Waterman, Adv. Appl. Math. 2:482 (1981), by the homology alignment algorithm of Needleman & Wunsch, J. Mol. Biol. 48:443 (1970), by the search for

similarity method of Pearson & Lipman, Proc. Nat'l. Acad. Sci. USA 85:2444 (1988), by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, Wis.), or by visual inspection (see generally Ausubel et al., *infra*).

[00151] One example of an algorithm that is suitable for determining percent sequence identity and sequence similarity is the BLAST algorithm, which is described in Altschul et al., J. Mol. Biol. 215:403-410 (1990). Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information.

[00152] As used herein the term “non-stop or read-through” is a mutation causing the removal of the natural stop codon.

[00153] As used herein the term “epitope” is the specific portion of an antigen typically bound by an antibody or T cell receptor.

[00154] As used herein the term “immunogenic” is the ability to elicit an immune response, e.g., via T cells, B cells, or both.

[00155] As used herein the term “HLA binding affinity” “MHC binding affinity” means affinity of binding between a specific antigen and a specific MHC allele.

[00156] As used herein the term “bait” is a nucleic acid probe used to enrich a specific sequence of DNA or RNA from a sample.

[00157] As used herein the term “variant” is a difference between a subject's nucleic acids and the reference human genome used as a control.

[00158] As used herein the term “variant call” is an algorithmic determination of the presence of a variant, typically from sequencing.

[00159] As used herein the term “polymorphism” is a germline variant, i.e., a variant found in all DNA-bearing cells of an individual.

[00160] As used herein the term “somatic variant” is a variant arising in non-germline cells of an individual.

[00161] As used herein the term “allele” is a version of a gene or a version of a genetic sequence or a version of a protein.

[00162] As used herein the term “HLA type” is the complement of HLA gene alleles.

[00163] As used herein the term “nonsense-mediated decay” or “NMD” is a degradation of an mRNA by a cell due to a premature stop codon.

[00164] As used herein the term “exome” is a subset of the genome that codes for proteins. An exome can be the collective exons of a genome.

[00165] As used herein the term “logistic regression” is a regression model for binary data from statistics where the logit of the probability that the dependent variable is equal to one is modeled as a linear function of the dependent variables.

[00166] As used herein the term “neural network” is a machine learning model for classification or regression consisting of multiple layers of linear transformations followed by element-wise nonlinearities typically trained via stochastic gradient descent and back-propagation.

[00167] As used herein the term “proteome” is the set of all proteins expressed and/or translated by a cell, group of cells, or individual.

[00168] As used herein the term “peptidome” is the set of all peptides presented by MHC-I or MHC-II on the cell surface. The peptidome may refer to a property of a cell or a collection of cells (e.g., the infectious disease peptidome, meaning the union of the peptidomes of all cells that are infected by the infectious disease).

[00169] As used herein the term “ELISpot” means Enzyme-linked immunosorbent spot assay – which is a common method for monitoring immune responses in humans and animals.

[00170] As used herein the term “dextramers” is a dextran-based peptide-MHC multimers used for antigen-specific T-cell staining in flow cytometry.

[00171] As used herein the term “tolerance or immune tolerance” is a state of immune non-responsiveness to one or more antigens, e.g. self-antigens.

[00172] As used herein the term “central tolerance” is a tolerance affected in the thymus, either by deleting self-reactive T-cell clones or by promoting self-reactive T-cell clones to differentiate into immunosuppressive regulatory T-cells (Tregs).

[00173] As used herein the term “peripheral tolerance” is a tolerance affected in the periphery by downregulating or anergizing self-reactive T-cells that survive central tolerance or promoting these T cells to differentiate into Tregs.

[00174] The term “sample” can include a single cell or multiple cells or fragments of cells or an aliquot of body fluid, taken from a subject, by means including venipuncture, excretion, ejaculation, massage, biopsy, needle aspirate, lavage sample, scraping, surgical incision, or intervention or other means known in the art.

[00175] The term “subject” encompasses a cell, tissue, or organism, human or non-human, whether in vivo, ex vivo, or in vitro, male or female. The term subject is inclusive of mammals including humans.

[00176] The term “mammal” encompasses both humans and non-humans and includes but is not limited to humans, non-human primates, canines, felines, murines, bovines, equines, and porcines.

[00177] The term “clinical factor” refers to a measure of a condition of a subject, e.g., disease activity or severity. “Clinical factor” encompasses all markers of a subject’s health status, including non-sample markers, and/or other characteristics of a subject, such as, without limitation, age and gender. A clinical factor can be a score, a value, or a set of values that can be obtained from evaluation of a sample (or population of samples) from a subject or a subject under a determined condition. A clinical factor can also be predicted by markers and/or other parameters such as gene expression surrogates. Clinical factors can include infection type, infection sub-type, and smoking history.

[00178] The term “antigen-encoding nucleic acid sequences derived from an infection” refers to nucleic acid sequences obtained from infected cells or a infectious disease organism, e.g. via RT-PCR; or sequence data obtained by sequencing the infected cell or infectious disease organism and then synthesizing the nucleic acid sequences using the sequencing data, e.g., via various synthetic or PCR-based methods known in the art. Derived sequences can include nucleic acid sequence variants, such as sequence-optimized nucleic acid sequence variants (e.g., codon-optimized and/or otherwise optimized for expression), that encode the same polypeptide sequence as the corresponding native infectious disease organism nucleic acid sequence. Derived sequences can include nucleic acid sequence variants that encode a modified infectious disease organism polypeptide sequence having one or more (e.g., 1, 2, 3, 4, or 5) mutations relative to a native infectious disease organism polypeptide sequence. For example, a modified polypeptide sequence can have one or more missense mutations relative to the native polypeptide sequence of an infectious disease organism protein.

[00179] The term “alphavirus” refers to members of the family *Togaviridae*, and are positive-sense single-stranded RNA viruses. Alphaviruses are typically classified as either Old World, such as Sindbis, Ross River, Mayaro, Chikungunya, and Semliki Forest viruses, or New World, such as eastern equine encephalitis, Aura, Fort Morgan, or Venezuelan equine encephalitis and its derivative strain TC-83. Alphaviruses are typically self-replicating RNA viruses.

[00180] The term “alphavirus backbone” refers to minimal sequence(s) of an alphavirus that allow for self-replication of the viral genome. Minimal sequences can include conserved sequences for nonstructural protein-mediated amplification, a nonstructural protein 1 (nsP1)

gene, a nsP2 gene, a nsP3 gene, a nsP4 gene, and a polyA sequence, as well as sequences for expression of subgenomic viral RNA including a 26S promoter element.

[00181] The term “sequences for nonstructural protein-mediated amplification” includes alphavirus conserved sequence elements (CSE) well known to those in the art. CSEs include, but are not limited to, an alphavirus 5' UTR, a 51-nt CSE, a 24-nt CSE, or other 26S subgenomic promoter sequence, a 19-nt CSE, and an alphavirus 3' UTR.

[00182] The term “RNA polymerase” includes polymerases that catalyze the production of RNA polynucleotides from a DNA template. RNA polymerases include, but are not limited to, bacteriophage derived polymerases including T3, T7, and SP6.

[00183] The term “lipid” includes hydrophobic and/or amphiphilic molecules. Lipids can be cationic, anionic, or neutral. Lipids can be synthetic or naturally derived, and in some instances biodegradable. Lipids can include cholesterol, phospholipids, lipid conjugates including, but not limited to, polyethyleneglycol (PEG) conjugates (PEGylated lipids), waxes, oils, glycerides, fats, and fat-soluble vitamins. Lipids can also include dilinoleylmethyl- 4-dimethylaminobutyrate (MC3) and MC3-like molecules.

[00184] The term “lipid nanoparticle” or “LNP” includes vesicle like structures formed using a lipid containing membrane surrounding an aqueous interior, also referred to as liposomes. Lipid nanoparticles includes lipid-based compositions with a solid lipid core stabilized by a surfactant. The core lipids can be fatty acids, acylglycerols, waxes, and mixtures of these surfactants. Biological membrane lipids such as phospholipids, sphingomyelins, bile salts (sodium taurocholate), and sterols (cholesterol) can be utilized as stabilizers. Lipid nanoparticles can be formed using defined ratios of different lipid molecules, including, but not limited to, defined ratios of one or more cationic, anionic, or neutral lipids. Lipid nanoparticles can encapsulate molecules within an outer-membrane shell and subsequently can be contacted with target cells to deliver the encapsulated molecules to the host cell cytosol. Lipid nanoparticles can be modified or functionalized with non-lipid molecules, including on their surface. Lipid nanoparticles can be single-layered (unilamellar) or multi-layered (multilamellar). Lipid nanoparticles can be complexed with nucleic acid. Unilamellar lipid nanoparticles can be complexed with nucleic acid, wherein the nucleic acid is in the aqueous interior. Multilamellar lipid nanoparticles can be complexed with nucleic acid, wherein the nucleic acid is in the aqueous interior, or to form or sandwiched between

[00185] Abbreviations: MHC: major histocompatibility complex; HLA: human leukocyte antigen, or the human MHC gene locus; NGS: next-generation sequencing; PPV: positive

predictive value; TSNA: tumor-specific neoantigen; FFPE: formalin-fixed, paraffin-embedded; NMD: nonsense-mediated decay; NSCLC: non-small-cell lung cancer; DC: dendritic cell.

[00186] It should be noted that, as used in the specification and the appended claims, the singular forms “a,” “an,” and “the” include plural referents unless the context clearly dictates otherwise.

[00187] Unless specifically stated or otherwise apparent from context, as used herein the term “about” is understood as within a range of normal tolerance in the art, for example within 2 standard deviations of the mean. About can be understood as within 10%, 9%, 8%, 7%, 6%, 5%, 4%, 3%, 2%, 1%, 0.5%, 0.1%, 0.05%, or 0.01% of the stated value. Unless otherwise clear from context, all numerical values provided herein are modified by the term about.

[00188] Any terms not directly defined herein shall be understood to have the meanings commonly associated with them as understood within the art of the invention. Certain terms are discussed herein to provide additional guidance to the practitioner in describing the compositions, devices, methods and the like of aspects of the invention, and how to make or use them. It will be appreciated that the same thing may be said in more than one way. Consequently, alternative language and synonyms may be used for any one or more of the terms discussed herein. No significance is to be placed upon whether or not a term is elaborated or discussed herein. Some synonyms or substitutable methods, materials and the like are provided. Recital of one or a few synonyms or equivalents does not exclude use of other synonyms or equivalents, unless it is explicitly stated. Use of examples, including examples of terms, is for illustrative purposes only and does not limit the scope and meaning of the aspects of the invention herein.

[00189] All references, issued patents and patent applications cited within the body of the specification are hereby incorporated by reference in their entirety, for all purposes.

II. Antigen Identification

[00190] Research methods for NGS analysis of tumor and normal exome and transcriptomes have been described and applied in the antigen identification space.^{6,14,15} Certain optimizations for greater sensitivity and specificity for antigen identification in the clinical setting can be considered. These optimizations can be grouped into two areas, those related to laboratory processes and those related to the NGS data analysis. The research methods described can also be applied to identification of antigens in other settings, such as identification of identifying antigens from an infectious disease organism, an infection in a subject, or an infected cell of a

subject. Examples of optimizations are known to those skilled in the art, for example the methods described in more detail in US Pat No. 10,055,540, US Application Pub. No. US20200010849A1, international patent application publications WO/2018/195357 and WO/2018/208856, US App. No. 16/606,577, and international patent application PCT/US2020/021508, each herein incorporated by reference, in their entirety, for all purposes.

[00191] Methods for identifying antigens (*e.g.*, antigens derived from an infectious disease organism) include identifying antigens that are likely to be presented on a cell surface (*e.g.*, presented by MHC on an infected cell or an immune cell, including professional antigen presenting cells such as dendritic cells), and/or are likely to be immunogenic. As an example, one such method may comprise the steps of: obtaining at least one of exome, transcriptome or whole genome nucleotide sequencing and/or expression data from an infected cell or an infectious disease organism, wherein the nucleotide sequencing data and/or expression data from the an infected cell of the subject, wherein the infectious disease organism nucleotide sequencing and/or expression data is used to obtain data representing peptide sequences of each of a set of antigens (*e.g.*, antigens derived from an infectious disease organism); inputting the peptide sequence of each antigen into one or more presentation models to generate a set of numerical likelihoods that each of the antigens is presented by one or more MHC alleles on the cell surface of an infected cell of the subject or cells present in the subject, the set of numerical likelihoods having been identified at least based on received mass spectrometry data; and selecting a subset of the set of antigens based on the set of numerical likelihoods to generate a set of selected antigens.

[00192] The presentation model can comprise a statistical regression or a machine learning (*e.g.*, deep learning) model trained on a set of reference data (also referred to as a training data set) comprising a set of corresponding labels, wherein the set of reference data is obtained from each of a plurality of distinct subjects where optionally some subjects can have an infection, and wherein the set of reference data comprises at least one of: data representing exome nucleotide sequences from infected tissue, data representing exome nucleotide sequences from normal tissue, data representing transcriptome nucleotide sequences from infected tissue, data representing proteome sequences from infected tissue, and data representing MHC peptidome sequences from infected tissue, and data representing MHC peptidome sequences from normal tissue. The reference data can further comprise mass spectrometry data, sequencing data, RNA sequencing data, expression profiling data, and proteomics data for single-allele cell lines engineered to express a predetermined MHC allele that are subsequently exposed to synthetic

protein, normal and tumor human cell lines, and fresh and frozen primary samples, and T cell assays (e.g., ELISpot). In certain aspects, the set of reference data includes each form of reference data.

[00193] The presentation model can comprise a set of features derived at least in part from the set of reference data, and wherein the set of features comprises at least one of allele dependent-features and allele-independent features. In certain aspects each feature is included.

[00194] Methods for identifying shared antigens also include generating an output for constructing a personalized cancer vaccine by identifying one or more antigens from one or more cells of a subject that are likely to be presented on a surface of infected cells. As an example, one such method may comprise the steps of: obtaining at least one of exome, transcriptome, or whole genome nucleotide sequencing and/or expression data from the infected cells and normal cells of the subject, wherein the nucleotide sequencing and/or expression data is used to obtain data representing peptide sequences of each of a set of antigens identified by comparing the nucleotide sequencing and/or expression data from the infected cells and the nucleotide sequencing and/or expression data from the normal cells, peptide sequence identified from the normal cells of the subject; encoding the peptide sequences of each of the antigens into a corresponding numerical vector, each numerical vector including information regarding a plurality of amino acids that make up the peptide sequence and a set of positions of the amino acids in the peptide sequence; inputting the numerical vectors, using a computer processor, into a deep learning presentation model to generate a set of presentation likelihoods for the set of antigens, each presentation likelihood in the set representing the likelihood that a corresponding antigen is presented by one or more class II MHC alleles on the surface of the infected cells of the subject, the deep learning presentation model; selecting a subset of the set of antigens based on the set of presentation likelihoods to generate a set of selected antigens; and generating the output for constructing the personalized cancer vaccine based on the set of selected antigens.

[00195] Specific methods for identifying antigens (e.g., infectious disease organism derived antigens) are known to those skilled in the art, for example the methods described in more detail in international patent application publications WO/2017/106638, WO/2018/195357, and WO/2018/208856, each herein incorporated by reference, in their entirety, for all purposes.

[00196] A method of treating a subject having an infection is disclosed herein, comprising performing the steps of any of the antigen identification methods described herein, and further

comprising obtaining a infectious disease vaccine comprising the set of selected antigens, and administering the infectious disease vaccine to the subject.

[00197] A method disclosed herein can also include identifying one or more T cells that are antigen-specific for at least one of the antigens in the subset. In some embodiments, the identification comprises co-culturing the one or more T cells with one or more of the antigens in the subset under conditions that expand the one or more antigen-specific T cells. In further embodiments, the identification comprises contacting the one or more T cells with a tetramer comprising one or more of the antigens in the subset under conditions that allow binding between the T cell and the tetramer. In even further embodiments, the method disclosed herein can also include identifying one or more T cell receptors (TCR) of the one or more identified T cells. In certain embodiments, identifying the one or more T cell receptors comprises sequencing the T cell receptor sequences of the one or more identified T cells. The method disclosed herein can further comprise genetically engineering a plurality of T cells to express at least one of the one or more identified T cell receptors; culturing the plurality of T cells under conditions that expand the plurality of T cells; and infusing the expanded T cells into the subject. In some embodiments, genetically engineering the plurality of T cells to express at least one of the one or more identified T cell receptors comprises cloning the T cell receptor sequences of the one or more identified T cells into an expression vector; and transfecting each of the plurality of T cells with the expression vector. In some embodiments, the method disclosed herein further comprises culturing the one or more identified T cells under conditions that expand the one or more identified T cells; and infusing the expanded T cells into the subject.

[00198] Also disclosed herein is an isolated T cell that is antigen-specific for at least one selected antigen in the subset.

[00199] Also disclosed herein is a methods for manufacturing an infectious disease vaccine, comprising the steps of: obtaining at least one of exome, transcriptome or whole genome infectious disease organism nucleotide sequencing and/or expression data from the infected cell of the subject, wherein the infectious disease organism nucleotide sequencing and/or expression data is used to obtain data representing peptide sequences of each of a set of antigens (*e.g.*, where peptides are derived from any polypeptide known to or have been found to have altered expression in a infected cell or infected tissue in comparison to a normal cell or tissue); inputting the peptide sequence of each antigen into one or more presentation models to generate a set of numerical likelihoods that each of the antigens is presented by one or more

MHC alleles on the cell surface of the infected cell of the subject, the set of numerical likelihoods having been identified at least based on received mass spectrometry data; and selecting a subset of the set of antigens based on the set of numerical likelihoods to generate a set of selected antigens; and producing or having produced a infectious disease vaccine comprising the set of selected antigens.

[00200] Also disclosed herein is an infectious disease vaccine including a set of selected antigens selected by performing the method comprising the steps of: obtaining at least one of exome, transcriptome or whole genome infectious disease organism nucleotide sequencing and/or expression data from the infected cell of the subject, wherein the infectious disease organism nucleotide sequencing and/or expression data is used to obtain data representing peptide sequences of each of a set of antigens, and wherein the peptide sequence of each antigen (*e.g.*, derived from any polypeptide known to or have been found to have altered expression in a infected cell or infected tissue in comparison to a normal cell or tissue); inputting the peptide sequence of each antigen into one or more presentation models to generate a set of numerical likelihoods that each of the antigens is presented by one or more MHC alleles on the cell surface of the infected cell of the subject, the set of numerical likelihoods having been identified at least based on received mass spectrometry data; and selecting a subset of the set of antigens based on the set of numerical likelihoods to generate a set of selected antigens; and producing or having produced an infectious disease vaccine comprising the set of selected antigens.

[00201] The vaccine may include one or more of a nucleotide sequence, a polypeptide sequence, RNA, DNA, a cell, a plasmid, or a vector.

[00202] The vaccine may include one or more antigens presented on the infected cell surface.

[00203] The infectious disease vaccine may include one or more antigens that is immunogenic in the subject.

[00204] The infectious disease vaccine may not include one or more antigens that induce an autoimmune response against normal tissue in the subject.

[00205] The infectious disease vaccine may include an adjuvant.

[00206] The infectious disease vaccine may include an excipient.

[00207] A method disclosed herein may also include selecting antigens that have an increased likelihood of being presented on the infected cell surface relative to unselected antigens based on the presentation model.

[00208] A method disclosed herein may also include selecting antigens that have an increased likelihood of being capable of inducing a infectious disease organism-specific immune response in the subject relative to unselected antigens based on the presentation model.

[00209] A method disclosed herein may also include selecting antigens that have an increased likelihood of being capable of being presented to naïve T cells by professional antigen presenting cells (APCs) relative to unselected antigens based on the presentation model, optionally wherein the APC is a dendritic cell (DC).

[00210] A method disclosed herein may also include selecting antigens that have a decreased likelihood of being subject to inhibition via central or peripheral tolerance relative to unselected antigens based on the presentation model.

[00211] A method disclosed herein may also include selecting antigens that have a decreased likelihood of being capable of inducing an autoimmune response to normal tissue in the subject relative to unselected antigens based on the presentation model.

[00212] The exome or transcriptome nucleotide sequencing and/or expression data may be obtained by performing sequencing on the infected tissue.

[00213] The sequencing may be next generation sequencing (NGS) or any massively parallel sequencing approach.

[00214] The set of numerical likelihoods may be further identified by at least MHC-allele interacting features comprising at least one of: the predicted affinity with which the MHC allele and the antigen encoded peptide bind; the predicted stability of the antigen encoded peptide-MHC complex; the sequence and length of the antigen encoded peptide; the probability of presentation of antigen encoded peptides with similar sequence in cells from other individuals expressing the particular MHC allele as assessed by mass-spectrometry proteomics or other means; the expression levels of the particular MHC allele in the subject in question (e.g. as measured by RNA-seq or mass spectrometry); the overall antigen encoded peptide-sequence-independent probability of presentation by the particular MHC allele in other distinct subjects who express the particular MHC allele; the overall antigen encoded peptide-sequence-independent probability of presentation by MHC alleles in the same family of molecules (e.g., HLA-A, HLA-B, HLA-C, HLA-DQ, HLA-DR, HLA-DP) in other distinct subjects.

[00215] The set of numerical likelihoods are further identified by at least MHC-allele noninteracting features comprising at least one of: the C- and N-terminal sequences flanking the antigen encoded peptide within its source protein sequence; the presence of protease cleavage motifs in the antigen encoded peptide, optionally weighted according to the

expression of corresponding proteases in the infected cells (as measured by RNA-seq or mass spectrometry); the turnover rate of the source protein as measured in the appropriate cell type; the length of the source protein, optionally considering the specific splice variants (“isoforms”) most highly expressed in the infected cells as measured by RNA-seq or proteome mass spectrometry, or as predicted from the annotation of germline or somatic splicing mutations detected in DNA or RNA sequence data; the level of expression of the proteasome, immunoproteasome, thymoproteasome, or other proteases in the infected cells (which may be measured by RNA-seq, proteome mass spectrometry, or immunohistochemistry); the expression of the source gene of the antigen encoded peptide (e.g., as measured by RNA-seq or mass spectrometry); the typical tissue-specific expression of the source gene of the antigen encoded peptide during various stages of the cell cycle; a comprehensive catalog of features of the source protein and/or its domains as can be found in e.g. uniProt or PDB <http://www.rcsb.org/pdb/home/home.do>; features describing the properties of the domain of the source protein containing the peptide, for example: secondary or tertiary structure (e.g., alpha helix vs beta sheet); alternative splicing; the probability of presentation of peptides from the source protein of the antigen encoded peptide in question in other distinct subjects; the probability that the peptide will not be detected or over-represented by mass spectrometry due to technical biases; the expression of various gene modules/pathways as measured by RNASeq (which need not contain the source protein of the peptide) that are informative about the state of the infected cells, stroma, or infected tissue; the copy number of the source gene of the antigen encoded peptide in the infected cells; the probability that the peptide binds to the TAP or the measured or predicted binding affinity of the peptide to the TAP; the expression level of TAP in the infected cells (which may be measured by RNA-seq, proteome mass spectrometry, immunohistochemistry); presence or absence of tumor mutations, including, but not limited to: driver mutations in known cancer driver genes such as EGFR, KRAS, ALK, RET, ROS1, TP53, CDKN2A, CDKN2B, NTRK1, NTRK2, NTRK3, and in genes encoding the proteins involved in the antigen presentation machinery (e.g., B2M, HLA-A, HLA-B, HLA-C, TAP-1, TAP-2, TAPBP, CALR, CNX, ERP57, HLA-DM, HLA-DMA, HLA-DMB, HLA-DO, HLA-DOA, HLA-DOB, HLA-DP, HLA-DPA1, HLA-DPB1, HLA-DQ, HLA-DQA1, HLA-DQA2, HLA-DQB1, HLA-DQB2, HLA-DR, HLA-DRA, HLA-DRB1, HLA-DRB3, HLA-DRB4, HLA-DRB5 or any of the genes coding for components of the proteasome or immunoproteasome). Peptides whose presentation relies on a component of the antigen-presentation machinery that is subject to loss-of-function mutation in the tumor have reduced

probability of presentation; presence or absence of functional germline polymorphisms, including, but not limited to: in genes encoding the proteins involved in the antigen presentation machinery (e.g., B2M, HLA-A, HLA-B, HLA-C, TAP-1, TAP-2, TAPBP, CALR, CNX, ERP57, HLA-DM, HLA-DMA, HLA-DMB, HLA-DO, HLA-DOA, HLA-DOB, HLA-DP, HLA-DPA1, HLA-DPB1, HLA-DQ, HLA-DQA1, HLA-DQA2, HLA-DQB1, HLA-DQB2, HLA-DR, HLA-DRA, HLA-DRB1, HLA-DRB3, HLA-DRB4, HLA-DRB5 or any of the genes coding for components of the proteasome or immunoproteasome); infection type (e.g., a pathogen infection, a viral infection, a bacterial infection, a fungal infection, and a parasitic infection); clinical infection subtype (e.g., an HIV infection, an HPV infection, a SARS infection, a SARS-CoV-2 infection, an Ebola infection, a HBV infection, an influenza infection, a HCV infection, a CMV infection, a Chikungunya virus infection, a RSV infection, a Dengue virus infection, a orthomyxoviridae family virus infection, and a tuberculosis infection); smoking history; the typical expression of the source gene of the peptide in the relevant infection type or clinical subtype.

[00216] A method disclosed herein may also include obtaining a infectious disease vaccine comprising the set of selected antigens (e.g., infectious disease organism derived antigens) or a subset thereof, optionally further comprising administering the infectious disease vaccine to the subject.

[00217] At least one of the antigens (e.g., infectious disease organism derived antigens) in the set of selected antigens, when in polypeptide form, may include at least one of: a binding affinity with MHC with an IC₅₀ value of less than 1000nM, for MHC Class I polypeptides a length of 8-15, 8, 9, 10, 11, 12, 13, 14, or 15 amino acids, for MHC Class II polypeptides a length of 6-30, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, or 30 amino acids, presence of sequence motifs within or near the polypeptide in the parent protein sequence promoting proteasome cleavage, and presence of sequence motifs promoting TAP transport. For MHC Class II, presence of sequence motifs within or near the peptide promoting cleavage by extracellular or lysosomal proteases (e.g., cathepsins) or HLA-DM catalyzed HLA binding.

[00218]

[00219] Disclosed herein is are methods for identifying one or more antigens (e.g., infectious disease organism derived antigens) that are likely to be presented on a cell surface of an infected cell, comprising executing the steps of: receiving mass spectrometry data comprising data associated with a plurality of isolated peptides eluted from major

histocompatibility complex (MHC) derived from a plurality of fresh or frozen samples; obtaining a training data set by at least identifying a set of training peptide sequences present in the samples and presented on one or more MHC alleles associated with each training peptide sequence; obtaining a set of training protein sequences based on the training peptide sequences; and training a set of numerical parameters of a presentation model using the training protein sequences and the training peptide sequences, the presentation model providing a plurality of numerical likelihoods that peptide sequences from the infected cell are presented by one or more MHC alleles on the infected cell surface.

[00220] The presentation model may represent dependence between: presence of a pair of a particular one of the MHC alleles and a particular amino acid at a particular position of a peptide sequence; and likelihood of presentation on the infected cell surface, by the particular one of the MHC alleles of the pair, of such a peptide sequence comprising the particular amino acid at the particular position.

[00221] A method disclosed herein can also include selecting a subset of antigens (e.g., infectious disease organism derived antigens), wherein the subset of antigens is selected because each has an increased likelihood that it is presented on the cell surface of the infected cell relative to one or more distinct antigens.

[00222] A method disclosed herein can also include selecting a subset of antigens (e.g., infectious disease organism derived antigens), wherein the subset of antigens is selected because each has an increased likelihood that it is capable of inducing a disease-specific immune response in the subject relative to one or more distinct antigens.

[00223] A method disclosed herein can also include selecting a subset of antigens (e.g., infectious disease organism derived antigens), wherein the subset of antigens is selected because each has an increased likelihood that it is capable of being presented to naïve T cells by professional antigen presenting cells (APCs) relative to one or more distinct antigens, optionally wherein the APC is a dendritic cell (DC).

[00224] A method disclosed herein can also include selecting a subset of antigens (e.g., infectious disease organism derived antigens), wherein the subset of antigens is selected because each has a decreased likelihood that it is subject to inhibition via central or peripheral tolerance relative to one or more distinct antigens.

[00225] A method disclosed herein can also include selecting a subset of antigens (e.g., infectious disease organism derived antigens), wherein the subset of antigens is selected

because each has a decreased likelihood that it is capable of inducing an autoimmune response to normal tissue in the subject relative to one or more distinct antigens.

[00226] A method disclosed herein can also include selecting a subset of antigens (e.g., infectious disease organism derived antigens), wherein the subset of antigens is selected because each has a decreased likelihood that it will be differentially post-translationally modified in infected cells versus APCs, optionally wherein the APC is a dendritic cell (DC).

[00227] The practice of the methods herein will employ, unless otherwise indicated, conventional methods of protein chemistry, biochemistry, recombinant DNA techniques and pharmacology, within the skill of the art. Such techniques are explained fully in the literature. See, e.g., T.E. Creighton, *Proteins: Structures and Molecular Properties* (W.H. Freeman and Company, 1993); A.L. Lehninger, *Biochemistry* (Worth Publishers, Inc., current addition); Sambrook, et al., *Molecular Cloning: A Laboratory Manual* (2nd Edition, 1989); *Methods In Enzymology* (S. Colowick and N. Kaplan eds., Academic Press, Inc.); *Remington's Pharmaceutical Sciences*, 18th Edition (Easton, Pennsylvania: Mack Publishing Company, 1990); Carey and Sundberg *Advanced Organic Chemistry 3rd Ed.* (Plenum Press) Vols A and B(1992).

III. Antigens

[00228] Antigens can include nucleotides or polypeptides. For example, an antigen can be an RNA sequence that encodes for a polypeptide sequence. Antigens useful in vaccines can therefore include nucleotide sequences or polypeptide sequences. Antigens can be derived from nucleotide sequences or polypeptide sequences of an infectious disease organism. Polypeptide sequences of an infectious disease organism include, but are not limited to, a pathogen-derived peptide, a virus-derived peptide, a bacteria-derived peptide, a fungus-derived peptide, and/or a parasite-derived peptide. Infectious disease organism include, but are not limited to, Severe acute respiratory syndrome-related coronavirus (SARS), severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), Ebola, HIV, Hepatitis B virus (HBV), influenza, Hepatitis C virus (HCV), Human papillomavirus (HPV), Cytomegalovirus (CMV), Chikungunya virus, Respiratory syncytial virus (RSV), Dengue virus, a orthomyxoviridae family virus, and tuberculosis.

[00229] Disclosed herein are isolated peptides that comprise infectious disease organism specific antigens or epitopes identified by the methods disclosed herein, peptides that comprise known infectious disease organism specific antigens or epitopes, and mutant polypeptides or

fragments thereof identified by methods disclosed herein. Antigen peptides can be described in the context of their coding sequence where an antigen includes the nucleotide sequence (e.g., DNA or RNA) that codes for the related polypeptide sequence.

[00230] Antigens that can be incorporated into a vaccine (*e.g.*, encoded in a cassette) include immunogens which are useful to immunize a human or non-human animal against viruses, such as pathogenic viruses which infect human and non-human vertebrates. Antigens may be selected from a variety of viral families. Example of desirable viral families against which an immune response would be desirable include, the picornavirus family, which includes the genera rhinoviruses, which are responsible for about 50% of cases of the common cold; the genera enteroviruses, which include polioviruses, coxsackieviruses, echoviruses, and human enteroviruses such as hepatitis A virus; and the genera aphthoviruses, which are responsible for foot and mouth diseases, primarily in non-human animals. Within the picornavirus family of viruses, target antigens include the VP1, VP2, VP3, VP4, and VPG. Another viral family includes the calcivirus family, which encompasses the Norwalk group of viruses, which are an important causative agent of epidemic gastroenteritis. Still another viral family desirable for use in targeting antigens for inducing immune responses in humans and non-human animals is the togavirus family, which includes the genera alphavirus, which include Sindbis viruses, Ross River virus, and Venezuelan, Eastern & Western Equine encephalitis, and rubivirus, including Rubella virus. The Flaviviridae family includes dengue, yellow fever, Japanese encephalitis, St. Louis encephalitis and tick borne encephalitis viruses. Other target antigens may be generated from the Hepatitis C or the coronavirus family, which includes a number of non-human viruses such as infectious bronchitis virus (poultry), porcine transmissible gastroenteric virus (pig), porcine hemagglutinating encephalomyelitis virus (pig), feline infectious peritonitis virus (cats), feline enteric coronavirus (cat), canine coronavirus (dog), and human respiratory coronaviruses, which may cause the common cold and/or non-A, B or C hepatitis. Within the coronavirus family, target antigens include the E1 (also called M or matrix protein), E2 (also called S or Spike protein), E3 (also called HE or hemagglutinin-esterase) glycoprotein (not present in all coronaviruses), or N (nucleocapsid). Still other antigens may be targeted against the rhabdovirus family, which includes the genera vesiculovirus (*e.g.*, Vesicular Stomatitis Virus), and the general lyssavirus (*e.g.*, rabies). Within the rhabdovirus family, suitable antigens may be derived from the G protein or the N protein. The family filoviridae, which includes hemorrhagic fever viruses such as Marburg and Ebola virus, may be a suitable source of antigens. The paramyxovirus family includes parainfluenza Virus Type 1,

parainfluenza Virus Type 3, bovine parainfluenza Virus Type 3, rubulavirus (mumps virus), parainfluenza Virus Type 2, parainfluenza virus Type 4, Newcastle disease virus (chickens), rinderpest, morbillivirus, which includes measles and canine distemper, and pneumovirus, which includes respiratory syncytial virus (e.g., the glyco-(G) protein and the fusion (F) protein, for which sequences are available from GenBank). Influenza virus is classified within the family orthomyxovirus and can be suitable source of antigens (e.g., the HA protein, the N1 protein). The bunyavirus family includes the genera bunyavirus (California encephalitis, La Crosse), phlebovirus (Rift Valley Fever), hantavirus (pneumonia is a hantavirus fever virus), nairovirus (Nairobi sheep disease) and various unassigned bunyaviruses. The arenavirus family provides a source of antigens against LCM and Lassa fever virus. The reovirus family includes the genera reovirus, rotavirus (which causes acute gastroenteritis in children), orbiviruses, and cultivirus (Colorado Tick fever, Lebombo (humans), equine encephalosis, blue tongue). The retrovirus family includes the sub-family oncovirinae which encompasses such human and veterinary diseases as feline leukemia virus, HTLV I and HTLV II, lentivirinae (which includes human immunodeficiency virus (HIV), simian immunodeficiency virus (SIV), feline immunodeficiency virus (FIV), equine infectious anemia virus, and spumavirinae). Among the lentiviruses, many suitable antigens have been described and can readily be selected. Examples of suitable HIV and SIV antigens include, without limitation the gag, pol, Vif, Vpx, VPR, Env, Tat, Nef, and Rev proteins, as well as various fragments thereof. For example, suitable fragments of the Env protein may include any of its subunits such as the gp120, gp160, gp41, or smaller fragments thereof, e.g., of at least about 8 amino acids in length. Similarly, fragments of the tat protein may be selected. [See, U.S. Pat. No. 5,891,994 and U.S. Pat. No. 6,193,981.] See, also, the HIV and SIV proteins described in D. H. Barouch et al, *J. Virol.*, 75(5):2462-2467 (March 2001), and R. R. Amara, et al, *Science*, 292:69-74 (6 Apr. 2001). In another example, the HIV and/or SIV immunogenic proteins or peptides may be used to form fusion proteins or other immunogenic molecules. See, e.g., the HIV-1 Tat and/or Nef fusion proteins and immunization regimens described in WO 01/54719, published Aug. 2, 2001, and WO 99/16884, published Apr. 8, 1999. The invention is not limited to the HIV and/or SIV immunogenic proteins or peptides described herein. In addition, a variety of modifications to these proteins have been described or could readily be made by one of skill in the art. See, e.g., the modified gag protein that is described in U.S. Pat. No. 5,972,596. Further, any desired HIV and/or SIV immunogens may be delivered alone or in combination. Such combinations may include expression from a single vector or from multiple vectors. The papovavirus family

includes the sub-family polyomaviruses (BKU and JCU viruses) and the sub-family papillomavirus (associated with cancers or malignant progression of papilloma). The adenovirus family includes viruses (EX, AD7, ARD, O.B.) which cause respiratory disease and/or enteritis. The parvovirus family feline parvovirus (feline enteritis), feline panleucopeniavirus, canine parvovirus, and porcine parvovirus. The herpesvirus family includes the sub-family alphaherpesvirinae, which encompasses the genera simplexvirus (HSV1, HSV2), varicellovirus (pseudorabies, varicella zoster) and the sub-family betaherpesvirinae, which includes the genera cytomegalovirus (Human CMV), muromegalovirus) and the sub-family gammaherpesvirinae, which includes the genera lymphocryptovirus, EBV (Burkitts lymphoma), infectious rhinotracheitis, Marek's disease virus, and rhadinovirus. The poxvirus family includes the sub-family chordopoxvirinae, which encompasses the genera orthopoxvirus (Variola (Smallpox) and Vaccinia (Cowpox)), parapoxvirus, avipoxvirus, capripoxvirus, leporipoxvirus, suipoxvirus, and the sub-family entomopoxvirinae. The hepadnavirus family includes the Hepatitis B virus. One unclassified virus which may be suitable source of antigens is the Hepatitis delta virus. Still other viral sources may include avian infectious bursal disease virus and porcine respiratory and reproductive syndrome virus. The alphavirus family includes equine arteritis virus and various Encephalitis viruses.

[00231] Antigens that can be incorporated into a vaccine (*e.g.*, encoded in a cassette) also include immunogens which are useful to immunize a human or non-human animal against pathogens including bacteria, fungi, parasitic microorganisms or multicellular parasites which infect human and non-human vertebrates. Examples of bacterial pathogens include pathogenic gram-positive cocci include pneumococci; staphylococci; and streptococci. Pathogenic gram-negative cocci include meningococcus; gonococcus. Pathogenic enteric gram-negative bacilli include enterobacteriaceae; *pseudomonas*, acinetobacteria and *eikenella*; melioidosis; *salmonella*; *shigella*; *haemophilus* (*Haemophilus influenzae*, *Haemophilus somnus*); *moraxella*; *H. ducreyi* (which causes chancroid); *brucella*; *Francisella tularensis* (which causes tularemia); *yersinia (pasteurella)*; *streptobacillus moniliformis* and *spirillum*. Gram-positive bacilli include *listeria monocytogenes*; *erysipelotheix rhusiopathiae*; *Corynebacterium diphtheria* (diphtheria); cholera; *B. anthracis* (anthrax); donovanosis (granuloma inguinale); and bartonellosis. Diseases caused by pathogenic anaerobic bacteria include tetanus; botulism; other clostridia; tuberculosis; leprosy; and other mycobacteria. Examples of specific bacterium species are, without

limitation, *Streptococcus pneumoniae*, *Streptococcus pyogenes*, *Streptococcus agalactiae*, *Streptococcus faecalis*, *Moraxella catarrhalis*, *Helicobacter pylori*, *Neisseria meningitidis*, *Neisseria gonorrhoeae*, *Chlamydia trachomatis*, *Chlamydia pneumoniae*, *Chlamydia psittaci*, *Bordetella pertussis*, *Salmonella typhi*, *Salmonella typhimurium*, *Salmonella choleraesuis*, *Escherichia coli*, *Shigella*, *Vibrio cholerae*, *Corynebacterium diphtheriae*, *Mycobacterium tuberculosis*, *Mycobacterium avium*, *Mycobacterium intracellulare* complex, *Proteus mirabilis*, *Proteus vulgaris*, *Staphylococcus aureus*, *Clostridium tetani*, *Leptospira interrogans*, *Borrelia burgdorferi*, *Pasteurella haemolytica*, *Pasteurella multocida*, *Actinobacillus pleuropneumoniae* and *Mycoplasma gallisepticum*. Pathogenic spirochetal diseases include syphilis; treponematoses: yaws, pinta and endemic syphilis; and leptospirosis. Other infections caused by higher pathogen bacteria and pathogenic fungi include actinomycosis; nocardiosis; cryptococcosis (*Cryptococcus*), blastomycosis (*Blastomyces*), histoplasmosis (*Histoplasma*) and coccidioidomycosis (*Coccidioides*); candidiasis (*Candida*), aspergillosis (*Aspergillus*), and mucormycosis; sporotrichosis; paracoccidioidomycosis, petriellidiosis, torulopsosis, mycetoma and chromomycosis; and dermatophytosis. Rickettsial infections include Typhus fever, Rocky Mountain spotted fever, Q fever, and Rickettsialpox. Examples of mycoplasma and chlamydial infections include: *mycoplasma pneumoniae*; lymphogranuloma venereum; psittacosis; and perinatal chlamydial infections. Pathogenic eukaryotes encompass pathogenic protozoans and helminths and infections produced thereby include: amebiasis; malaria; leishmaniasis (e.g., caused by *Leishmania major*); trypanosomiasis; toxoplasmosis (e.g., caused by *Toxoplasma gondii*); *Pneumocystis carinii*; Trichans; *Toxoplasma gondii*; babesiosis; giardiasis (e.g., caused by *Giardia*); trichinosis (e.g., caused by *Trichomonas*); filariasis; schistosomiasis (e.g., caused by *Schistosoma*); nematodes; trematodes or flukes; and cestode (tapeworm) infections. Other parasitic infections may be caused by *Ascaris*, *Trichuris*, *Cryptosporidium*, and *Pneumocystis carinii*, among others.

[00232] Also disclosed herein are peptides derived from any polypeptide associated with an infectious disease organism, an infection in a subject, or an infected cell of a subject. Antigens can be derived from nucleic acid sequences or polypeptide sequences of an infectious disease organism. Polypeptide sequences of an infectious disease organism include, but are not limited to, a pathogen-derived peptide, a virus-derived peptide, a bacteria-derived peptide, a fungus-derived peptide, and/or a parasite-derived peptide. Infectious disease organism include, but are not limited to, Severe acute respiratory syndrome-related coronavirus (SARS), severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), Ebola, HIV, Hepatitis B virus (HBV),

influenza, Hepatitis C virus (HCV), Human papillomavirus (HPV), Cytomegalovirus (CMV), Chikungunya virus, Respiratory syncytial virus (RSV), Dengue virus, a orthomyxoviridae family virus, and tuberculosis.

[00233] Antigens can be selected that are predicted to be presented on the cell surface of a cell, such as an infected cell or an immune cell, including professional antigen presenting cells such as dendritic cells. Antigens can be selected that are predicted to be immunogenic.

[00234] One or more polypeptides encoded by an antigen nucleotide sequence can comprise at least one of: a binding affinity with MHC with an IC50 value of less than 1000nM, for MHC Class I peptides a length of 8-15, 8, 9, 10, 11, 12, 13, 14, or 15 amino acids, presence of sequence motifs within or near the peptide promoting proteasome cleavage, and presence or sequence motifs promoting TAP transport. For MHC Class II peptides a length 6-30, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, or 30 amino acids, presence of sequence motifs within or near the peptide promoting cleavage by extracellular or lysosomal proteases (e.g., cathepsins) or HLA-DM catalyzed HLA binding.

[00235] One or more antigens can be presented on the surface of an infected cell.

[00236] One or more antigens can be immunogenic in a subject having or suspected to have an infection, e.g., capable of eliciting a T cell response and/or a B cell response in the subject. One or more antigens can be immunogenic in a subject at risk of an infection infection, e.g., capable of eliciting a T cell response and/or a B cell response in the subject that provides immunological protection (*i.e.*, immunity) against the infection, *e.g.*, such as stimulating the production of memory T cells, memory B cells, and/or antibodies specific to the infection.

[00237] One or more antigens can be capable of eliciting a B cell response, such as the production of antibodies that recognize the one or more antigens (*e.g.*, antibodies that recognize an infectious disease antigens). Antibodies can recognize linear polypeptide sequences or recognize secondary and tertiary structures. Accordingly, B cell antigens can include linear polypeptide sequences or polypeptides having secondary and tertiary structures, including, but not limited to, full-length proteins, protein subunits, protein domains, or any polypeptide sequence known or predicted to have secondary and tertiary structures. In general, antigens capable of eliciting a B cell response to an infection are antigens found on the surface of an infectious disease organism.

[00238] One or more antigens can include a combination of antigens capable of eliciting a T cell response (*e.g.*, peptides including predicted T cell epitope sequences) and distinct antigens

capable of eliciting a T cell response (*e.g.*, full-length proteins, protein subunits, protein domains).

[00239] One or more antigens that induce an autoimmune response in a subject can be excluded from consideration in the context of vaccine generation for a subject.

[00240] The size of at least one antigenic peptide molecule (*e.g.*, an epitope sequence) can comprise, but is not limited to, about 5, about 6, about 7, about 8, about 9, about 10, about 11, about 12, about 13, about 14, about 15, about 16, about 17, about 18, about 19, about 20, about 21, about 22, about 23, about 24, about 25, about 26, about 27, about 28, about 29, about 30, about 31, about 32, about 33, about 34, about 35, about 36, about 37, about 38, about 39, about 40, about 41, about 42, about 43, about 44, about 45, about 46, about 47, about 48, about 49, about 50, about 60, about 70, about 80, about 90, about 100, about 110, about 120 or greater amino molecule residues, and any range derivable therein. In specific embodiments the antigenic peptide molecules are equal to or less than 50 amino acids.

[00241] Antigenic peptides and polypeptides can be: for MHC Class I 15 residues or less in length and usually consist of between about 8 and about 11 residues, particularly 9 or 10 residues; for MHC Class II, 6-30 residues, inclusive.

[00242] If desirable, a longer peptide can be designed in several ways. In one case, when presentation likelihoods of peptides on HLA alleles are predicted or known, a longer peptide could consist of either: (1) individual presented peptides with an extensions of 2-5 amino acids toward the N- and C-terminus of each corresponding gene product; (2) a concatenation of some or all of the presented peptides with extended sequences for each. In another case, when sequencing reveals a long (>10 residues) epitope sequence present, a longer peptide would consist of: (3) the entire stretch of novel infectious disease-specific amino acids -- thus bypassing the need for computational or in vitro test-based selection of the strongest HLA-presented shorter peptide. In both cases, use of a longer peptide allows endogenous processing by patient cells and may lead to more effective antigen presentation and induction of T cell responses. Longer peptides can also a full-length protein, a protein subunit, a protein domain, and combinations thereof of a peptide expressed in an infectious disease organism.

[00243] Antigenic peptides and polypeptides can be presented on an HLA protein. In some aspects antigenic peptides and polypeptides are presented on an HLA protein with greater affinity than a wild-type peptide. In some aspects, an antigenic peptide or polypeptide can have an IC₅₀ of at least less than 5000 nM, at least less than 1000 nM, at least less than 500 nM, at

least less than 250 nM, at least less than 200 nM, at least less than 150 nM, at least less than 100 nM, at least less than 50 nM or less.

[00244] In some aspects, antigenic peptides and polypeptides do not induce an autoimmune response and/or invoke immunological tolerance when administered to a subject.

[00245] Also provided are compositions comprising at least two or more antigenic peptides. In some embodiments the composition contains at least two distinct peptides. At least two distinct peptides can be derived from the same polypeptide. By distinct polypeptides is meant that the peptide vary by length, amino acid sequence, or both. The peptides can be derived from any polypeptide known to or suspected to be associated with an infectious disease organism, or peptides derived from any polypeptide known to or have been found to have altered expression in an infected cell in comparison to a normal cell or tissue (*e.g.*, an infectious disease polynucleotide or polypeptide, including infectious disease polynucleotides or polypeptides with expression restricted to a host cell).

[00246] Antigenic peptides and polypeptides having a desired activity or property can be modified to provide certain desired attributes, *e.g.*, improved pharmacological characteristics, while increasing or at least retaining substantially all of the biological activity of the unmodified peptide to bind the desired MHC molecule and activate the appropriate T cell. For instance, antigenic peptide and polypeptides can be subject to various changes, such as substitutions, either conservative or non-conservative, where such changes might provide for certain advantages in their use, such as improved MHC binding, stability or presentation. By conservative substitutions is meant replacing an amino acid residue with another which is biologically and/or chemically similar, *e.g.*, one hydrophobic residue for another, or one polar residue for another. The substitutions include combinations such as Gly, Ala; Val, Ile, Leu, Met; Asp, Glu; Asn, Gln; Ser, Thr; Lys, Arg; and Phe, Tyr. The effect of single amino acid substitutions may also be probed using D-amino acids. Such modifications can be made using well known peptide synthesis procedures, as described in *e.g.*, Merrifield, *Science* 232:341-347 (1986), Barany & Merrifield, *The Peptides*, Gross & Meienhofer, eds. (N.Y., Academic Press), pp. 1-284 (1979); and Stewart & Young, *Solid Phase Peptide Synthesis*, (Rockford, Ill., Pierce), 2d Ed. (1984).

[00247] Modifications of peptides and polypeptides with various amino acid mimetics or unnatural amino acids can be particularly useful in increasing the stability of the peptide and polypeptide *in vivo*. Stability can be assayed in a number of ways. For instance, peptidases and various biological media, such as human plasma and serum, have been used to test stability.

See, e.g., Verhoef et al., *Eur. J. Drug Metab Pharmacokin.* 11:291-302 (1986). Half-life of the peptides can be conveniently determined using a 25% human serum (v/v) assay. The protocol is generally as follows. Pooled human serum (Type AB, non-heat inactivated) is delipidated by centrifugation before use. The serum is then diluted to 25% with RPMI tissue culture media and used to test peptide stability. At predetermined time intervals a small amount of reaction solution is removed and added to either 6% aqueous trichloroacetic acid or ethanol. The cloudy reaction sample is cooled (4 degrees C) for 15 minutes and then spun to pellet the precipitated serum proteins. The presence of the peptides is then determined by reversed-phase HPLC using stability-specific chromatography conditions.

[00248] The peptides and polypeptides can be modified to provide desired attributes other than improved serum half-life. For instance, the ability of the peptides to induce CTL activity can be enhanced by linkage to a sequence which contains at least one epitope that is capable of inducing a T helper cell response. Immunogenic peptides/T helper conjugates can be linked by a spacer molecule. The spacer is typically comprised of relatively small, neutral molecules, such as amino acids or amino acid mimetics, which are substantially uncharged under physiological conditions. The spacers are typically selected from, e.g., Ala, Gly, or other neutral spacers of nonpolar amino acids or neutral polar amino acids. It will be understood that the optionally present spacer need not be comprised of the same residues and thus can be a hetero- or homo-oligomer. When present, the spacer will usually be at least one or two residues, more usually three to six residues. Alternatively, the peptide can be linked to the T helper peptide without a spacer.

[00249] A antigenic peptide can be linked to the T helper peptide either directly or via a spacer either at the amino or carboxy terminus of the peptide. The amino terminus of either the antigenic peptide or the T helper peptide can be acylated. Exemplary T helper peptides include tetanus toxoid 830-843, influenza 307-319, malaria circumsporozoite 382-398 and 378-389.

[00250] Proteins or peptides can be made by any technique known to those of skill in the art, including the expression of proteins, polypeptides or peptides through standard molecular biological techniques, the isolation of proteins or peptides from natural sources, or the chemical synthesis of proteins or peptides. The nucleotide and protein, polypeptide and peptide sequences corresponding to various genes have been previously disclosed, and can be found at computerized databases known to those of ordinary skill in the art. One such database is the National Center for Biotechnology Information's Genbank and GenPept databases located at the National Institutes of Health website. The coding regions for known genes can be amplified

and/or expressed using the techniques disclosed herein or as would be known to those of ordinary skill in the art. Alternatively, various commercial preparations of proteins, polypeptides and peptides are known to those of skill in the art.

[00251] In a further aspect an antigen includes a nucleic acid (e.g. polynucleotide) that encodes an antigenic peptide or portion thereof. The polynucleotide can be, e.g., DNA, cDNA, PNA, CNA, RNA (e.g., mRNA), either single- and/or double-stranded, or native or stabilized forms of polynucleotides, such as, e.g., polynucleotides with a phosphorothiate backbone, or combinations thereof and it may or may not contain introns. The polynucleotide sequence encoding an antigen can be sequence-optimized to improve expression, such as through improving transcription, translation, post-transcriptional processing, and/or RNA stability. For example, polynucleotide sequence encoding an antigen can be codon-optimized. “Codon-optimization” herein refers to replacing infrequently used codons, with respect to codon bias of a given organism, with frequently used synonymous codons. Polynucleotide sequences can be optimized to improve post-transcriptional processing, for example optimized to reduce unintended splicing, such as through removal of splicing motifs (e.g., canonical and/or cryptic/non-canonical splice donor, branch, and/or acceptor sequences) and/or introduction of exogenous splicing motifs (e.g., splice donor, branch, and/or acceptor sequences) to bias favored splicing events. Exogenous intron sequences include, but are not limited to, those derived from SV40 (e.g., an SV40 mini-intron) and derived from immunoglobulins (e.g., human β -globin gene). Exogenous intron sequences can be incorporated between a promoter/enhancer sequence and the antigen(s) sequence. Exogenous intron sequences for use in expression vectors are described in more detail in Callendret et al. (Virology. 2007 Jul 5; 363(2): 288–302), herein incorporated by reference for all purposes. Polynucleotide sequences can be optimized to improve transcript stability, for example through removal of RNA instability motifs (e.g., AU-rich elements and 3' UTR motifs) and/or repetitive nucleotide sequences. Polynucleotide sequences can be optimized to improve accurate transcription, for example through removal of cryptic transcriptional initiators and/or terminators. Polynucleotide sequences can be optimized to improve translation and translational accuracy, for example through removal of cryptic AUG start codons, premature polyA sequences, and/or secondary structure motifs. Polynucleotide sequences can be optimized to improve nuclear export of transcripts, such as through addition of a Constitutive Transport Element (CTE), RNA Transport Element (RTE), or Woodchuck Posttranscriptional Regulatory Element (WPRE). Nuclear export signals for use in expression vectors are described in more detail in Callendret

et al. (Virology. 2007 Jul 5; 363(2): 288–302), herein incorporated by reference for all purposes. Polynucleotide sequences can be optimized with respect to GC content, for example to reflect the average GC content of a given organism. Sequence optimization can balance one or more sequence properties, such as transcription, translation, post-transcriptional processing, and/or RNA stability. Sequence optimization can generate an optimal sequence balancing each of transcription, translation, post-transcriptional processing, and RNA stability. Sequence optimization algorithms are known to those of skill in the art, such as GeneArt (Thermo Fisher), Codon Optimization Tool (IDT), Cool Tool (University of Singapore), SGI-DNA (La Jolla California). One or more regions of an antigen-encoding protein can be sequence-optimized separately.

[00252] A still further aspect provides an expression vector capable of expressing a polypeptide or portion thereof. Expression vectors for different cell types are well known in the art and can be selected without undue experimentation. Generally, DNA is inserted into an expression vector, such as a plasmid, in proper orientation and correct reading frame for expression. If necessary, DNA can be linked to the appropriate transcriptional and translational regulatory control nucleotide sequences recognized by the desired host, although such controls are generally available in the expression vector. The vector is then introduced into the host through standard techniques. Guidance can be found e.g. in Sambrook et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

IV. Vaccine Compositions

[00253] Also disclosed herein is an immunogenic composition, e.g., a vaccine composition, capable of raising a specific immune response, e.g., a infectious disease organism-specific immune response. Vaccine compositions typically comprise one or a plurality of antigens, e.g., selected using a method described herein or selected from a pathogen-derived peptide, a virus-derived peptide, a bacteria-derived peptide, a fungus-derived peptide, and/or a parasite-derived peptide. Vaccine compositions can also be referred to as vaccines.

[00254] A vaccine can contain between 1 and 30 peptides, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, or 30 different peptides, 6, 7, 8, 9, 10 11, 12, 13, or 14 different peptides, or 12, 13 or 14 different peptides. Peptides can include post-translational modifications. A vaccine can contain between 1 and 100 or more nucleotide sequences, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48,

49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100 or more different nucleotide sequences, 6, 7, 8, 9, 10 11, 12, 13, or 14 different nucleotide sequences, or 12, 13 or 14 different nucleotide sequences. A vaccine can contain between 1 and 30 antigen sequences, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100 or more different antigen sequences, 6, 7, 8, 9, 10 11, 12, 13, or 14 different antigen sequences, or 12, 13 or 14 different antigen sequences.

[00255] A vaccine can contain between 1 and 30 antigen-encoding nucleic acid sequences, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100 or more different antigen-encoding nucleic acid sequences, 6, 7, 8, 9, 10 11, 12, 13, or 14 different antigen-encoding nucleic acid sequences, or 12, 13 or 14 different antigen-encoding nucleic acid sequences. Antigen-encoding nucleic acid sequences can refer to the antigen encoding portion of an antigen “cassette.” Features of an antigen cassette are described herein. An antigen-encoding nucleic acid sequence can contain one or more epitope-encoding nucleic acid sequences (*e.g.*, an antigen-encoding nucleic acid sequence encoding concatenated T cell epitopes).

[00256] A vaccine can contain between 1 and 30 distinct epitope-encoding nucleic acid sequences, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100 or more distinct epitope-encoding nucleic acid sequences, 6, 7, 8, 9, 10 11, 12, 13, or 14 distinct epitope-encoding nucleic acid sequences, or 12, 13 or 14 distinct epitope-encoding nucleic acid sequences. Epitope-encoding nucleic acid sequences can refer to sequences for individual epitope sequences, such as each of the T cell epitopes in an antigen-encoding nucleic acid sequence encoding concatenated T cell epitopes.

[00257] A vaccine can contain at least two repeats of an epitope-encoding nucleic acid sequence. A used herein, a “repeat” refers to two or more iterations of an identical nucleic acid epitope-encoding nucleic acid sequence (inclusive of the optional 5’ linker sequence and/or the optional 3’ linker sequences described herein) within an antigen-encoding nucleic acid sequence. In one example, the antigen-encoding nucleic acid sequence portion of a cassette encodes at least two repeats of an epitope-encoding nucleic acid sequence. In further non-limiting examples, the antigen-encoding nucleic acid sequence portion of a cassette encodes more than one distinct epitope, and at least one of the distinct epitopes is encoded by at least two repeats of the nucleic acid sequence encoding the distinct epitope (*i.e.*, at least two distinct epitope-encoding nucleic acid sequences). In illustrative non-limiting examples, an antigen-encoding nucleic acid sequence encodes epitopes A, B, and C encoded by epitope-encoding nucleic acid sequences epitope-encoding sequence A (E_A), epitope-encoding sequence B (E_B), and epitope-encoding sequence C (E_C), and exemplary antigen-encoding nucleic acid sequences having repeats of at least one of the distinct epitopes are illustrated by, but is not limited to, the formulas below:

- Repeat of one distinct epitope (repeat of epitope A):

E_A-E_B-E_C-E_A; or

E_A-E_A-E_B-E_C

- Repeat of multiple distinct epitopes (repeats of epitopes A, B, and C):

E_A-E_B-E_C-E_A-E_B-E_C; or

E_A-E_A-E_B-E_B-E_C-E_C

- Multiple repeats of multiple distinct epitopes (repeats of epitopes A, B, and C):

E_A-E_B-E_C-E_A-E_B-E_C-E_A-E_B-E_C; or

E_A-E_A-E_A-E_B-E_B-E_B-E_C-E_C-E_C

[00258] The above examples are not limiting and the antigen-encoding nucleic acid sequences having repeats of at least one of the distinct epitopes can encode each of the distinct epitopes in any order or frequency. For example, the order and frequency can be a random arrangement of the distinct epitopes, *e.g.*, in an example with epitopes A, B, and C, by the formula E_A-E_B-E_C-E_C-E_A-E_B-E_A-E_C-E_A-E_C-E_C-E_B.

[00259] Also provided for herein is an antigen-encoding cassette, the antigen-encoding cassette having at least one antigen-encoding nucleic acid sequence described, from 5’ to 3’, by the formula:

$$(E_x-(E^n_n)_y)_z$$

where E represents a nucleotide sequence comprising at least one of the at least one distinct epitope-encoding nucleic acid sequences,

n represents the number of separate distinct epitope-encoding nucleic acid sequences and is any integer including 0,

E^N represents a nucleotide sequence comprising the separate distinct epitope-encoding nucleic acid sequence for each corresponding n,

for each iteration of z: $x = 0$ or 1 , $y = 0$ or 1 for each n, and at least one of x or $y = 1$, and

$z = 2$ or greater, wherein the antigen-encoding nucleic acid sequence comprises at least two iterations of E, a given E^N , or a combination thereof.

[00260] Each E or E^N can independently comprise any epitope-encoding nucleic acid sequence described herein (*e.g.*, a peptide encoding an infectious disease T cell epitope). For example, Each E or E^N can independently comprises a nucleotide sequence described, from 5' to 3', by the formula ($L5_b-N_c-L3_d$), where N comprises the distinct epitope-encoding nucleic acid sequence associated with each E or E^N , where $c = 1$, L5 comprises a 5' linker sequence, where $b = 0$ or 1 , and L3 comprises a 3' linker sequence, where $d = 0$ or 1 . Epitopes and linkers that can be used are further described herein.

[00261] Repeats of an epitope-encoding nucleic acid sequences (inclusive of optional 5' linker sequence and/or the optional 3' linker sequences) can be linearly linked directly to one another (*e.g.*, $E_A-E_A\ldots$ as illustrated above). Repeats of an epitope-encoding nucleic acid sequences can be separated by one or more additional nucleotides sequences. In general, repeats of an epitope-encoding nucleic acid sequences can be separated by any size nucleotide sequence applicable for the compositions described herein. In one example, repeats of an epitope-encoding nucleic acid sequences can be separated by a separate distinct epitope-encoding nucleic acid sequence (*e.g.*, $E_A-E_B-E_C-E_A\ldots$, as illustrated above). In examples where repeats are separated by a single separate distinct epitope-encoding nucleic acid sequence, and each epitope-encoding nucleic acid sequences (inclusive of optional 5' linker sequence and/or the optional 3' linker sequences) encodes a peptide 25 amino acids in length, the repeats can be separated by 75 nucleotides, such as in antigen-encoding nucleic acid represented by $E_A-E_B-E_A\ldots$, E_A is separated by 75 nucleotides. In an illustrative example, an antigen-encoding nucleic acid having the sequence

VTNTEMFVTAPDNLGYMYEVQWPGQTQPQIANCSVYDFFVWLHYYSVRDTVTNTEM
FVTAPDNLGYMYEVQWPGQTQPQIANCSVYDFFVWLHYYSVRDT encoding repeats of
25mer antigens Trp1 (VTNTEMFVTAPDNLGYMYEVQWPGQ) and Trp2

(TQPQIANCSVYDFFVWLHYYSVRDT), the repeats of Trp1 are separated by the 25mer Trp2 and thus the repeats of the Trp1 epitope-encoding nucleic acid sequences are separated the 75 nucleotide Trp2 epitope-encoding nucleic acid sequence. In examples where repeats are separated by 2, 3, 4, 5, 6, 7, 8, or 9 separate distinct epitope-encoding nucleic acid sequence, and each epitope-encoding nucleic acid sequences (inclusive of optional 5' linker sequence and/or the optional 3' linker sequences) encodes a peptide 25 amino acids in length, the repeats can be separated by 150, 225, 300, 375, 450, 525, 600, or 675 nucleotides, respectively.

[00262] In one embodiment, different peptides and/or polypeptides or nucleotide sequences encoding them are selected so that the peptides and/or polypeptides capable of associating with different MHC molecules, such as different MHC class I molecules and/or different MHC class II molecules. In some aspects, one vaccine composition comprises coding sequence for peptides and/or polypeptides capable of associating with the most frequently occurring MHC class I molecules and/or different MHC class II molecules. Hence, vaccine compositions can comprise different fragments capable of associating with at least 2 preferred, at least 3 preferred, or at least 4 preferred MHC class I molecules and/or different MHC class II molecules.

[00263] The vaccine composition can be capable of raising a specific cytotoxic T-cells response and/or a specific helper T-cell response.

[00264] The vaccine composition can be capable of raising a specific B-cell response (*e.g.*, an antibody response).

[00265] The vaccine composition can be capable of raising a specific cytotoxic T-cells response, a specific helper T-cell response, and/or a specific B-cell response (*e.g.*, an antibody response). The vaccine composition can be capable of raising a specific cytotoxic T-cells response, a specific helper T-cell response, and a specific B-cell response (*e.g.*, an antibody response).

[00266] A vaccine composition can further comprise an adjuvant and/or a carrier. Examples of useful adjuvants and carriers are given herein below. A composition can be associated with a carrier such as *e.g.* a protein or an antigen-presenting cell such as a dendritic cell (DC) capable of presenting the peptide to a T-cell.

[00267] Adjuvants are any substance whose admixture into a vaccine composition increases or otherwise modifies the immune response to an antigen. Carriers can be scaffold structures, for example a polypeptide or a polysaccharide, to which an antigen, is capable of being associated. Optionally, adjuvants are conjugated covalently or non-covalently.

[00268] The ability of an adjuvant to increase an immune response to an antigen is typically manifested by a significant or substantial increase in an immune-mediated reaction, or reduction in disease symptoms. For example, an increase in humoral immunity is typically manifested by a significant increase in the titer of antibodies raised to the antigen, and an increase in T-cell activity is typically manifested in increased cell proliferation, or cellular cytotoxicity, or cytokine secretion. An adjuvant may also alter an immune response, for example, by changing a primarily humoral or Th response into a primarily cellular, or Th response.

[00269] Suitable adjuvants include, but are not limited to 1018 ISS, alum, aluminium salts, Amplivax, AS15, BCG, CP-870,893, CpG7909, CyaA, dSLIM, GM-CSF, IC30, IC31, Imiquimod, ImuFact IMP321, IS Patch, ISS, ISCOMATRIX, JuvImmune, LipoVac, MF59, monophosphoryl lipid A, Montanide IMS 1312, Montanide ISA 206, Montanide ISA 50V, Montanide ISA-51, OK-432, OM-174, OM-197-MP-EC, ONTAK, PepTel vector system, PLG microparticles, resiquimod, SRL172, Virosomes and other Virus-like particles, YF-17D, VEGF trap, R848, beta-glucan, Pam3Cys, Aquila's QS21 stimulon (Aquila Biotech, Worcester, Mass., USA) which is derived from saponin, mycobacterial extracts and synthetic bacterial cell wall mimics, and other proprietary adjuvants such as Ribi's Detox. Quil or Superfos. Adjuvants such as incomplete Freund's or GM-CSF are useful. Several immunological adjuvants (e.g., MF59) specific for dendritic cells and their preparation have been described previously (Dupuis M, et al., Cell Immunol. 1998; 186(1):18-27; Allison A C; Dev Biol Stand. 1998; 92:3-11). Also cytokines can be used. Several cytokines have been directly linked to influencing dendritic cell migration to lymphoid tissues (e.g., TNF-alpha), accelerating the maturation of dendritic cells into efficient antigen-presenting cells for T-lymphocytes (e.g., GM-CSF, IL-1 and IL-4) (U.S. Pat. No. 5,849,589, specifically incorporated herein by reference in its entirety) and acting as immunoadjuvants (e.g., IL-12) (Gabrilovich D I, et al., J Immunother Emphasis Tumor Immunol. 1996 (6):414-418).

[00270] CpG immunostimulatory oligonucleotides have also been reported to enhance the effects of adjuvants in a vaccine setting. Other TLR binding molecules such as RNA binding TLR 7, TLR 8 and/or TLR 9 may also be used.

[00271] Other examples of useful adjuvants include, but are not limited to, chemically modified CpGs (e.g. CpR, Idera), Poly(I:C)(e.g. polyi:CI2U), non-CpG bacterial DNA or RNA as well as immunoactive small molecules and antibodies such as cyclophosphamide, sunitinib, bevacizumab, celebrex, NCX-4016, sildenafil, tadalafil, vardenafil, sorafenib, XL-999, CP-

547632, pazopanib, ZD2171, AZD2171, ipilimumab, tremelimumab, and SC58175, which may act therapeutically and/or as an adjuvant. The amounts and concentrations of adjuvants and additives can readily be determined by the skilled artisan without undue experimentation. Additional adjuvants include colony-stimulating factors, such as Granulocyte Macrophage Colony Stimulating Factor (GM-CSF, sargramostim).

[00272] A vaccine composition can comprise more than one different adjuvant.

Furthermore, a therapeutic composition can comprise any adjuvant substance including any of the above or combinations thereof. It is also contemplated that a vaccine and an adjuvant can be administered together or separately in any appropriate sequence.

[00273] A carrier (or excipient) can be present independently of an adjuvant. The function of a carrier can for example be to increase the molecular weight of in particular mutant to increase activity or immunogenicity, to confer stability, to increase the biological activity, or to increase serum half-life. Furthermore, a carrier can aid presenting peptides to T-cells. A carrier can be any suitable carrier known to the person skilled in the art, for example a protein or an antigen presenting cell. A carrier protein could be but is not limited to keyhole limpet hemocyanin, serum proteins such as transferrin, bovine serum albumin, human serum albumin, thyroglobulin or ovalbumin, immunoglobulins, or hormones, such as insulin or palmitic acid. For immunization of humans, the carrier is generally a physiologically acceptable carrier acceptable to humans and safe. However, tetanus toxoid and/or diphtheria toxoid are suitable carriers. Alternatively, the carrier can be dextrans for example sepharose.

[00274] Cytotoxic T-cells (CTLs) recognize an antigen in the form of a peptide bound to an MHC molecule rather than the intact foreign antigen itself. The MHC molecule itself is located at the cell surface of an antigen presenting cell. Thus, an activation of CTLs is possible if a trimeric complex of peptide antigen, MHC molecule, and APC is present. Correspondingly, it may enhance the immune response if not only the peptide is used for activation of CTLs, but if additionally APCs with the respective MHC molecule are added. Therefore, in some embodiments a vaccine composition additionally contains at least one antigen presenting cell.

[00275] Antigens can also be included in viral vector-based vaccine platforms, such as vaccinia, fowlpox, self-replicating alphavirus, marabavirus, adenovirus (*See, e.g., Tatsis et al., Adenoviruses, Molecular Therapy* (2004) 10, 616—629), or lentivirus, including but not limited to second, third or hybrid second/third generation lentivirus and recombinant lentivirus of any generation designed to target specific cell types or receptors (*See, e.g., Hu et al., Immunization Delivered by Lentiviral Vectors for Cancer and Infectious Diseases, Immunol*

Rev. (2011) 239(1): 45-61, Sakuma et al., Lentiviral vectors: basic to translational, *Biochem J.* (2012) 443(3):603-18, Cooper et al., Rescue of splicing-mediated intron loss maximizes expression in lentiviral vectors containing the human ubiquitin C promoter, *Nucl. Acids Res.* (2015) 43 (1): 682-690, Zufferey et al., Self-Inactivating Lentivirus Vector for Safe and Efficient In Vivo Gene Delivery, *J. Virol.* (1998) 72 (12): 9873-9880). Dependent on the packaging capacity of the above mentioned viral vector-based vaccine platforms, this approach can deliver one or more nucleotide sequences that encode one or more antigen peptides. The sequences may be flanked by non-mutated sequences, may be separated by linkers or may be preceded with one or more sequences targeting a subcellular compartment (See, e.g., Gros et al., Prospective identification of neoantigen-specific lymphocytes in the peripheral blood of melanoma patients, *Nat Med.* (2016) 22 (4):433-8, Stronen et al., Targeting of cancer neoantigens with donor-derived T cell receptor repertoires, *Science.* (2016) 352 (6291):1337-41, Lu et al., Efficient identification of mutated cancer antigens recognized by T cells associated with durable tumor regressions, *Clin Cancer Res.* (2014) 20(13):3401-10). Upon introduction into a host, infected cells express the antigens, and thereby elicit a host immune (e.g., CTL) response against the peptide(s). Vaccinia vectors and methods useful in immunization protocols are described in, e.g., U.S. Pat. No. 4,722,848. Another vector is BCG (Bacille Calmette Guerin). BCG vectors are described in Stover et al. (Nature 351:456-460 (1991)). A wide variety of other vaccine vectors useful for therapeutic administration or immunization of antigens, e.g., Salmonella typhi vectors, and the like will be apparent to those skilled in the art from the description herein.

IV.A. Antigen Cassette

[00276] The methods employed for the selection of one or more antigens, the cloning and construction of a “cassette” and its insertion into a viral vector are within the skill in the art given the teachings provided herein. By “antigen cassette” or “cassette” is meant the combination of a selected antigen or plurality of antigens (e.g., antigen-encoding nucleic acid sequences) and the other regulatory elements necessary to transcribe the antigen(s) and express the transcribed product. The selected antigen or plurality of antigens can refer to distinct epitope sequences, e.g., an antigen-encoding nucleic acid sequence in the cassette can encode an epitope-encoding nucleic acid sequence (or plurality of epitope-encoding nucleic acid sequences) such that the epitopes are transcribed and expressed. An antigen or plurality of antigens can be operatively linked to regulatory components in a manner which permits

transcription. Such components include conventional regulatory elements that can drive expression of the antigen(s) in a cell transfected with the viral vector. Thus the antigen cassette can also contain a selected promoter which is linked to the antigen(s) and located, with other, optional regulatory elements, within the selected viral sequences of the recombinant vector. Cassettes can include one or more antigens, such as one or more pathogen-derived peptides, a virus-derived peptides, bacteria-derived peptides, fungus-derived peptides, and/or parasite-derived peptides. A cassette can have one or more antigen-encoding nucleic acid sequences, such as a cassette containing multiple antigen-encoding nucleic acid sequences each independently operably linked to separate promoters and/or linked together using other multicistronic systems, such as 2A ribosome skipping sequence elements (*e.g.*, E2A, P2A, F2A, or T2A sequences) or Internal Ribosome Entry Site (IRES) sequence elements. A linker can also have a cleavage site, such as a TEV or furin cleavage site. Linkers with cleavage sites can be used in combination with other elements, such as those in a multicistronic system. In a non-limiting illustrative example, a furin protease cleavage site can be used in conjunction with a 2A ribosome skipping sequence element such that the furin protease cleavage site is configured to facilitate removal of the 2A sequence following translation. In a cassette containing more than one antigen-encoding nucleic acid sequences, each antigen-encoding nucleic acid sequence can contain one or more epitope-encoding nucleic acid sequences (*e.g.*, an antigen-encoding nucleic acid sequence encoding concatenated T cell epitopes).

[00277] Useful promoters can be constitutive promoters or regulated (inducible) promoters, which will enable control of the amount of antigen(s) to be expressed. For example, a desirable promoter is that of the cytomegalovirus immediate early promoter/enhancer [see, *e.g.*, Boshart et al, Cell, 41:521-530 (1985)]. Another desirable promoter includes the Rous sarcoma virus LTR promoter/enhancer. Still another promoter/enhancer sequence is the chicken cytoplasmic beta-actin promoter [T. A. Kost et al, Nucl. Acids Res., 11(23):8287 (1983)]. Other suitable or desirable promoters can be selected by one of skill in the art.

[00278] The antigen cassette can also include nucleic acid sequences heterologous to the viral vector sequences including sequences providing signals for efficient polyadenylation of the transcript (poly(A), poly-A or pA) and introns with functional splice donor and acceptor sites. A common poly-A sequence which is employed in the exemplary vectors of this invention is that derived from the papovavirus SV-40. The poly-A sequence generally can be inserted in the cassette following the antigen-based sequences and before the viral vector sequences. A common intron sequence can also be derived from SV-40, and is referred to as

the SV-40 T intron sequence. an antigen cassette can also contain such an intron, located between the promoter/enhancer sequence and the antigen(s). Selection of these and other common vector elements are conventional [see, e.g., Sambrook et al, "Molecular Cloning. A Laboratory Manual.", 2d edit., Cold Spring Harbor Laboratory, New York (1989) and references cited therein] and many such sequences are available from commercial and industrial sources as well as from Genbank.

[00279] A antigen cassette can have one or more antigens. For example, a given cassette can include 1-10, 1-20, 1-30, 10-20, 15-25, 15-20, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, or more antigens. Antigens can be linked directly to one another. Antigens can also be linked to one another with linkers. Antigens can be in any orientation relative to one another including N to C or C to N.

[00280] As described elsewhere, the antigen cassette can be located in the site of any selected deletion in the viral vector backbone, such as the site of the E1 gene region deletion or E3 gene region deletion of a ChAd-based vector or the deleted structural proteins of a VEE backbone, among others which may be selected.

[00281] The antigen cassette can be described using the following formula to describe the ordered sequence of each element, from 5' to 3':

$$(P_a-(L5_b-N_c-L3_d)_X)_Z-(P2_h-(G5_e-U_f)_Y)_W-G3_g$$

[00282] wherein P and P2 comprise promoter nucleotide sequences, N comprises an MHC class I epitope encoding nucleic acid sequence, L5 comprises a 5' linker sequence, L3 comprises a 3' linker sequence, G5 comprises a nucleic acid sequences encoding an amino acid linker, G3 comprises one of the at least one nucleic acid sequences encoding an amino acid linker, U comprises an MHC class II antigen-encoding nucleic acid sequence, where for each X the corresponding Nc is a epitope encoding nucleic acid sequence, where for each Y the corresponding Uf is a universal MHC class II epitope-encoding nucleic acid sequence, optionally wherein the at least one universal sequence comprises at least one of Tetanus toxoid and PADRE. The composition and ordered sequence can be further defined by selecting the number of elements present, for example where a = 0 or 1, where b = 0 or 1, where c = 1, where d = 0 or 1, where e = 0 or 1, where f = 1, where g = 0 or 1, where h = 0 or 1, X = 1 to 400, Y = 0, 1, 2, 3, 4 or 5, Z = 1 to 400, and W = 0, 1, 2, 3, 4 or 5.

[00283] In one example, elements present include where a = 0, b = 1, d = 1, e = 1, g = 1, h = 0, X = 10, Y = 2, Z = 1, and W = 1, describing where no additional promoter is present (e.g.,

only the promoter nucleotide sequence provided by the vector backbone, such as an RNA alphavirus backbone, is present), 10 MHC class I epitope are present, a 5' linker is present for each N, a 3' linker is present for each N, 2 MHC class II epitopes are present, a linker is present linking the two MHC class II epitopes, a linker is present linking the 5' end of the two MHC class II epitopes to the 3' linker of the final MHC class I epitope, and a linker is present linking the 3' end of the two MHC class II epitopes to the to the vectorbackbone. Examples of linking the 3' end of the antigen cassette to the vectorbackbone include linking directly to the 3' UTR elements provided by the vectorbackbone, such as a 3' 19-nt CSE. Examples of linking the 5' end of the antigen cassette to the vector backbone include linking directly to a promoter or 5' UTR element of the vector backbone, such as a 26S promoter sequence, an alphavirus 5' UTR, a 51-nt CSE, or a 24-nt CSE.

[00284] Other examples include: where a = 1 describing where a promoter other than the promoter nucleotide sequence provided by the vector backbone is present; where a = 1 and Z is greater than 1 where multiple promoters other than the promoter nucleotide sequence provided by the vector backbone are present each driving expression of 1 or more distinct MHC class I epitope encoding nucleic acid sequences; where h = 1 describing where a separate promoter is present to drive expression of the MHC class II epitope-encoding nucleic acid sequences; and where g = 0 describing the MHC class II epitope-encoding nucleic acid sequence, if present, is directly linked to the vector backbone.

[00285] Other examples include where each MHC class I epitope that is present can have a 5' linker, a 3' linker, neither, or both. In examples where more than one MHC class I epitope is present in the same antigen cassette, some MHC class I epitopes may have both a 5' linker and a 3' linker, while other MHC class I epitopes may have either a 5' linker, a 3' linker, or neither. In other examples where more than one MHC class I epitope is present in the same antigen cassette, some MHC class I epitopes may have either a 5' linker or a 3' linker, while other MHC class I epitopes may have either a 5' linker, a 3' linker, or neither.

[00286] Other examples include where each antigen that is present can have a 5' linker, a 3' linker, neither, or both. In examples where more than one antigen is present in the same antigen cassette, some antigens may have both a 5' linker and a 3' linker, while other antigens may have either a 5' linker, a 3' linker, or neither. In other examples where more than one antigen is present in the same antigen cassette, some antigens may have either a 5' linker or a 3' linker, while other antigens may have either a 5' linker, a 3' linker, or neither.

[00287] In examples where more than one MHC class II epitope is present in the same antigen cassette, some MHC class II epitopes may have both a 5' linker and a 3' linker, while other MHC class II epitopes may have either a 5' linker, a 3' linker, or neither. In other examples where more than one MHC class II epitope is present in the same antigen cassette, some MHC class II epitopes may have either a 5' linker or a 3' linker, while other MHC class II epitopes may have either a 5' linker, a 3' linker, or neither.

[00288] The promoter nucleotide sequences P and/or P2 can be the same as a promoter nucleotide sequence provided by the vector backbone, such as a RNA alphavirus backbone. For example, the promoter sequence provided by the vector backbone, Pn and P2, can each comprise a 26S subgenomic promoter or a CMV promoter. The promoter nucleotide sequences P and/or P2 can be different from the promoter nucleotide sequence provided by the vector backbone, as well as can be different from each other.

[00289] The 5' linker L5 can be a native sequence or a non-natural sequence. Non-natural sequence include, but are not limited to, AAY, RR, and DPP. The 3' linker L3 can also be a native sequence or a non-natural sequence. Additionally, L5 and L3 can both be native sequences, both be non-natural sequences, or one can be native and the other non-natural. For each X, the amino acid linkers can be 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100 or more amino acids in length. For each X, the amino acid linkers can be also be at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 11, at least 12, at least 13, at least 14, at least 15, at least 16, at least 17, at least 18, at least 19, at least 20, at least 21, at least 22, at least 23, at least 24, at least 25, at least 26, at least 27, at least 28, at least 29, or at least 30 amino acids in length.

[00290] The amino acid linker G5, for each Y, can be 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100 or more amino acids in length. For each Y, the amino acid linkers can be also be at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 11, at least 12, at least 13, at least 14, at least 15, at least 16, at least 17, at

least 18, at least 19, at least 20, at least 21, at least 22, at least 23, at least 24, at least 25, at least 26, at least 27, at least 28, at least 29, or at least 30 amino acids in length.

[00291] The amino acid linker G3 can be 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100 or more amino acids in length. G3 can be also be at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 11, at least 12, at least 13, at least 14, at least 15, at least 16, at least 17, at least 18, at least 19, at least 20, at least 21, at least 22, at least 23, at least 24, at least 25, at least 26, at least 27, at least 28, at least 29, or at least 30 amino acids in length.

[00292] For each X, each N can encode a MHC class I epitope, a MHC class II epitope, an epitope capable of stimulating a B cell response, or a combination thereof. For each X, N can encode a combination of a MHC class I epitope, a MHC class II epitope, and an epitope capable of stimulating a B cell response. For each X, N can encode a combination of a MHC class I epitope and a MHC class II epitope. For each X, N can encode a combination of a MHC class I epitope and an epitope capable of stimulating a B cell response. For each X, N can encode a combination of a MHC class II epitope and an epitope capable of stimulating a B cell response. For each X, each N can encode a MHC class I epitope 7-15 amino acids in length. For each X, each N can also encodes a MHC class I epitope 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, or 30 amino acids in length. For each X, each N can also encodes a MHC class I epitope at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 11, at least 12, at least 13, at least 14, at least 15, at least 16, at least 17, at least 18, at least 19, at least 20, at least 21, at least 22, at least 23, at least 24, at least 25, at least 26, at least 27, at least 28, at least 29, or at least 30 amino acids in length. For each X, each N can encode a MHC class II epitope. For each X, each N can encode an epitope capable of stimulating a B cell response.

[00293] The cassette encoding the one or more antigens can be 700 nucleotides or less. The cassette encoding the one or more antigens can be 700 nucleotides or less and encode 2 distinct epitope-encoding nucleic acid sequences (*e.g.*, encode 2 distinct SARS-CoV-2 derived nucleic acid sequence encoding an immunogenic polypeptide). The cassette encoding the one or more antigens can be 700 nucleotides or less and encode at least 2 distinct epitope-encoding nucleic acid sequences. The cassette encoding the one or more antigens can be 700 nucleotides or less

and encode 3 distinct epitope-encoding nucleic acid sequences. The cassette encoding the one or more antigens can be 700 nucleotides or less and encode at least 3 distinct epitope-encoding nucleic acid sequences. The cassette encoding the one or more antigens can be 700 nucleotides or less and include 1-10, 1-5, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more antigens.

[00294] The cassette encoding the one or more antigens can be between 375-700 nucleotides in length. The cassette encoding the one or more antigens can be between 375-700 nucleotides in length and encode 2 distinct epitope-encoding nucleic acid sequences. The cassette encoding the one or more antigens can be between 375-700 nucleotides in length and encode at least 2 distinct epitope-encoding nucleic acid sequences. The cassette encoding the one or more antigens can be between 375-700 nucleotides in length and encode 3 distinct epitope-encoding nucleic acid sequences. The cassette encoding the one or more antigens be between 375-700 nucleotides in length and encode at least 3 distinct epitope-encoding nucleic acid sequences. The cassette encoding the one or more antigens can be between 375-700 nucleotides in length and include 1-10, 1-5, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more antigens.

[00295] The cassette encoding the one or more antigens can be 600, 500, 400, 300, 200, or 100 nucleotides in length or less. The cassette encoding the one or more antigens can be 600, 500, 400, 300, 200, or 100 nucleotides in length or less and encode 2 distinct epitope-encoding nucleic acid sequences. The cassette encoding the one or more antigens can be 600, 500, 400, 300, 200, or 100 nucleotides in length or less and encode at least 2 distinct epitope-encoding nucleic acid sequences. The cassette encoding the one or more antigens can be 600, 500, 400, 300, 200, or 100 nucleotides in length or less and encode 3 distinct epitope-encoding nucleic acid sequences. The cassette encoding the one or more antigens can be 600, 500, 400, 300, 200, or 100 nucleotides in length or less and encode at least 3 distinct epitope-encoding nucleic acid sequences. The cassette encoding the one or more antigens can be 600, 500, 400, 300, 200, or 100 nucleotides in length or less and include 1-10, 1-5, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more antigens.

[00296] The cassette encoding the one or more antigens can be between 375-600, between 375-500, or between 375-400 nucleotides in length. The cassette encoding the one or more antigens can be between 375-600, between 375-500, or between 375-400 nucleotides in length and encode 2 distinct epitope-encoding nucleic acid sequences. The cassette encoding the one or more antigens can be between 375-600, between 375-500, or between 375-400 nucleotides in length and encode at least 2 distinct epitope-encoding nucleic acid sequences. The cassette encoding the one or more antigens can be between 375-600, between 375-500, or between 375-

400 nucleotides in length and encode 3 distinct epitope-encoding nucleic acid sequences. The cassette encoding the one or more antigens can be between 375-600, between 375-500, or between 375-400 nucleotides in length and encode at least 3 distinct epitope-encoding nucleic acid sequences. The cassette encoding the one or more antigens can be between 375-600, between 375-500, or between 375-400 nucleotides in length and include 1-10, 1-5, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more antigens.

IV.B. Immune Checkpoints

[00297] Vectors described herein, such as C68 vectors described herein or alphavirus vectors described herein, can comprise a nucleic acid which encodes at least one antigen and the same or a separate vector can comprise a nucleic acid which encodes at least one immune modulator (e.g., an antibody such as an scFv) which binds to and blocks the activity of an immune checkpoint molecule. Vectors can comprise an antigen cassette and one or more nucleic acid molecules encoding a checkpoint inhibitor.

[00298] Illustrative immune checkpoint molecules that can be targeted for blocking or inhibition include, but are not limited to, CTLA-4, 4-1BB (CD137), 4-1BBL (CD137L), PDL1, PDL2, PD1, B7-H3, B7-H4, BTLA, HVEM, TIM3, GAL9, LAG3, TIM3, B7H3, B7H4, VISTA, KIR, 2B4 (belongs to the CD2 family of molecules and is expressed on all NK, $\gamma\delta$, and memory CD8⁺ ($\alpha\beta$) T cells), CD160 (also referred to as BY55), and CGEN-15049. Immune checkpoint inhibitors include antibodies, or antigen binding fragments thereof, or other binding proteins, that bind to and block or inhibit the activity of one or more of CTLA-4, PDL1, PDL2, PD1, B7-H3, B7-H4, BTLA, HVEM, TIM3, GAL9, LAG3, TIM3, B7H3, B7H4, VISTA, KIR, 2B4, CD160, and CGEN-15049. Illustrative immune checkpoint inhibitors include Tremelimumab (CTLA-4 blocking antibody), anti-OX40, PD-L1 monoclonal Antibody (Anti-B7-H1; MEDI4736), ipilimumab, MK-3475 (PD-1 blocker), Nivolumab (anti-PD1 antibody), CT-011 (anti-PD1 antibody), BY55 monoclonal antibody, AMP224 (anti-PDL1 antibody), BMS-936559 (anti-PDL1 antibody), MPLDL3280A (anti-PDL1 antibody), MSB0010718C (anti-PDL1 antibody) and Yervoy/ipilimumab (anti-CTLA-4 checkpoint inhibitor). Antibody-encoding sequences can be engineered into vectors such as C68 using ordinary skill in the art. An exemplary method is described in Fang et al., Stable antibody expression at therapeutic levels using the 2A peptide. *Nat Biotechnol.* 2005 May;23(5):584-90. Epub 2005 Apr 17; herein incorporated by reference for all purposes.

IV.C. Additional Considerations for Vaccine Design and Manufacture

IV.C.1. Determination of a Set of Peptides that Cover All Tumor Subclones

[00299] Truncal peptides, meaning those presented by all or most subclones, can be prioritized for inclusion into the vaccine.⁵³ Optionally, if there are no truncal peptides predicted to be presented and immunogenic with high probability, or if the number of truncal peptides predicted to be presented and immunogenic with high probability is small enough that additional non-truncal peptides can be included in the vaccine, then further peptides can be prioritized by estimating the number and identity of subclones and choosing peptides so as to maximize the number of subclones covered by the vaccine.⁵⁴

IV.C.2. Antigen Prioritization

[00300] After all of the above antigen filters are applied, more candidate antigens may still be available for vaccine inclusion than the vaccine technology can support. Additionally, uncertainty about various aspects of the antigen analysis may remain and tradeoffs may exist between different properties of candidate vaccine antigens. Thus, in place of predetermined filters at each step of the selection process, an integrated multi-dimensional model can be considered that places candidate antigens in a space with at least the following axes and optimizes selection using an integrative approach.

1. Risk of auto-immunity or tolerance (risk of germline) (lower risk of auto-immunity is typically preferred)
2. Probability of sequencing artifact (lower probability of artifact is typically preferred)
3. Probability of immunogenicity (higher probability of immunogenicity is typically preferred)
4. Probability of presentation (higher probability of presentation is typically preferred)
5. Gene expression (higher expression is typically preferred)
6. Coverage of HLA genes (larger number of HLA molecules involved in the presentation of a set of antigens may lower the probability that an infected cell will escape immune attack via downregulation or mutation of HLA molecules)
7. Coverage of HLA classes (covering both HLA-I and HLA-II may increase the probability of therapeutic response and decrease the probability of infectious disease escape)

[00301] Additionally, optionally, antigens can be deprioritized (e.g., excluded) from the vaccination if they are predicted to be presented by HLA alleles lost or inactivated in either all or part of the patient's infected cell. HLA allele loss can occur by either somatic mutation, loss of heterozygosity, or homozygous deletion of the locus. Methods for detection of HLA allele

somatic mutation are well known in the art, e.g. (Shukla et al., 2015). Methods for detection of somatic LOH and homozygous deletion (including for HLA locus) are likewise well described. (Carter et al., 2012; McGranahan et al., 2017; Van Loo et al., 2010). Antigens can also be deprioritized if mass-spectrometry data indicates a predicted antigen is not presented by a predicted HLA allele.

IV.D. Self-Amplifying RNA Vectors

[00302] In general, all self-amplifying RNA (SAM) vectors contain a self-amplifying backbone derived from a self-replicating virus. The term “self-amplifying backbone” refers to minimal sequence(s) of a self-replicating virus that allows for self-replication of the viral genome. For example, minimal sequences that allow for self-replication of an alphavirus can include conserved sequences for nonstructural protein-mediated amplification (*e.g.*, a nonstructural protein 1 (nsP1) gene, a nsP2 gene, a nsP3 gene, a nsP4 gene, and/or a polyA sequence). A self-amplifying backbone can also include sequences for expression of subgenomic viral RNA (*e.g.*, a 26S promoter element for an alphavirus). SAM vectors can be positive-sense RNA polynucleotides or negative-sense RNA polynucleotides, such as vectors with backbones derived from positive-sense or negative-sense self-replicating viruses. Self-replicating viruses include, but are not limited to, alphaviruses, flaviviruses (*e.g.*, Kunjin virus), measles viruses, and rhabdoviruses (*e.g.*, rabies virus and vesicular stomatitis virus). Examples of SAM vector systems derived from self-replicating viruses are described in Lundstrom (Molecules. 2018 Dec 13;23(12). pii: E3310. doi: 10.3390/molecules23123310), herein incorporated by reference for all purposes.

IV.D.1. Alphavirus Biology

[00303] Alphaviruses are members of the family *Togaviridae*, and are positive-sense single stranded RNA viruses. Members are typically classified as either Old World, such as Sindbis, Ross River, Mayaro, Chikungunya, and Semliki Forest viruses, or New World, such as eastern equine encephalitis, Aura, Fort Morgan, or Venezuelan equine encephalitis virus and its derivative strain TC-83 (Strauss Microbial Review 1994). A natural alphavirus genome is typically around 12kb in length, the first two-thirds of which contain genes encoding non-structural proteins (nsPs) that form RNA replication complexes for self-replication of the viral genome, and the last third of which contains a subgenomic expression cassette encoding structural proteins for virion production (Frolov RNA 2001).

[00304] A model lifecycle of an alphavirus involves several distinct steps (Strauss Microbial Review 1994, Jose Future Microbiol 2009). Following virus attachment to a host cell, the virion fuses with membranes within endocytic compartments resulting in the eventual release of genomic RNA into the cytosol. The genomic RNA, which is in a plus-strand orientation and comprises a 5' methylguanylate cap and 3' polyA tail, is translated to produce non-structural proteins nsP1-4 that form the replication complex. Early in infection, the plus-strand is then replicated by the complex into a minus-strand template. In the current model, the replication complex is further processed as infection progresses, with the resulting processed complex switching to transcription of the minus-strand into both full-length positive-strand genomic RNA, as well as the 26S subgenomic positive-strand RNA containing the structural genes. Several conserved sequence elements (CSEs) of alphavirus have been identified to potentially play a role in the various RNA replication steps including; a complement of the 5' UTR in the replication of plus-strand RNAs from a minus-strand template, a 51-nt CSE in the replication of minus-strand synthesis from the genomic template, a 24-nt CSE in the junction region between the nsPs and the 26S RNA in the transcription of the subgenomic RNA from the minus-strand, and a 3' 19-nt CSE in minus-strand synthesis from the plus-strand template.

[00305] Following the replication of the various RNA species, virus particles are then typically assembled in the natural lifecycle of the virus. The 26S RNA is translated and the resulting proteins further processed to produce the structural proteins including capsid protein, glycoproteins E1 and E2, and two small polypeptides E3 and 6K (Strauss 1994). Encapsidation of viral RNA occurs, with capsid proteins normally specific for only genomic RNA being packaged, followed by virion assembly and budding at the membrane surface.

IV.D.2. Alphavirus as a delivery vector

[00306] Alphaviruses (including alphavirus sequences, features, and other elements) can be used to generate alphavirus-based delivery vectors (also be referred to as alphavirus vectors, alphavirus viral vectors, alphavirus vaccine vectors, self-replicating RNA (srRNA) vectors, or self-amplifying RNA (samRNA) vectors). Alphaviruses have previously been engineered for use as expression vector systems (Pushko 1997, Rheme 2004). Alphaviruses offer several advantages, particularly in a vaccine setting where heterologous antigen expression can be desired. Due to its ability to self-replicate in the host cytosol, alphavirus vectors are generally able to produce high copy numbers of the expression cassette within a cell resulting in a high level of heterologous antigen production. Additionally, the vectors are generally transient, resulting in improved biosafety as well as reduced induction of immunological tolerance to the

vector. The public, in general, also lacks pre-existing immunity to alphavirus vectors as compared to other standard viral vectors, such as human adenovirus. Alphavirus based vectors also generally result in cytotoxic responses to infected cells. Cytotoxicity, to a certain degree, can be important in a vaccine setting to properly illicit an immune response to the heterologous antigen expressed. However, the degree of desired cytotoxicity can be a balancing act, and thus several attenuated alphaviruses have been developed, including the TC-83 strain of VEE. Thus, an example of an antigen expression vector described herein can utilize an alphavirus backbone that allows for a high level of antigen expression, elicits a robust immune response to antigen, does not elicit an immune response to the vector itself, and can be used in a safe manner. Furthermore, the antigen expression cassette can be designed to elicit different levels of an immune response through optimization of which alphavirus sequences the vector uses, including, but not limited to, sequences derived from VEE or its attenuated derivative TC-83.

[00307] Several expression vector design strategies have been engineered using alphavirus sequences (Pushko 1997). In one strategy, a alphavirus vector design includes inserting a second copy of the 26S promoter sequence elements downstream of the structural protein genes, followed by a heterologous gene (Frolov 1993). Thus, in addition to the natural non-structural and structural proteins, an additional subgenomic RNA is produced that expresses the heterologous protein. In this system, all the elements for production of infectious virions are present and, therefore, repeated rounds of infection of the expression vector in non-infected cells can occur.

[00308] Another expression vector design makes use of helper virus systems (Pushko 1997). In this strategy, the structural proteins are replaced by a heterologous gene. Thus, following self-replication of viral RNA mediated by still intact non-structural genes, the 26S subgenomic RNA provides for expression of the heterologous protein. Traditionally, additional vectors that expresses the structural proteins are then supplied *in trans*, such as by co-transfection of a cell line, to produce infectious virus. A system is described in detail in USPN 8,093,021, which is herein incorporated by reference in its entirety, for all purposes. The helper vector system provides the benefit of limiting the possibility of forming infectious particles and, therefore, improves biosafety. In addition, the helper vector system reduces the total vector length, potentially improving the replication and expression efficiency. Thus, an example of an antigen expression vector described herein can utilize an alphavirus backbone wherein the structural proteins are replaced by an antigen cassette, the resulting vector both reducing biosafety

concerns, while at the same time promoting efficient expression due to the reduction in overall expression vector size.

IV.D.3. Self-Amplifying Virus production *in vitro*

[00309] A convenient technique well-known in the art for RNA production is *in vitro* transcription (IVT). In this technique, a DNA template of the desired vector is first produced by techniques well-known to those in the art, including standard molecular biology techniques such as cloning, restriction digestion, ligation, gene synthesis, and polymerase chain reaction (PCR).

[00310] The DNA template contains a RNA polymerase promoter at the 5' end of the sequence desired to be transcribed into RNA (*e.g.*, SAM). Promoters include, but are not limited to, bacteriophage polymerase promoters such as T3, T7, K11, or SP6. Depending on the specific RNA polymerase promoter sequence chosen, additional 5' nucleotides can transcribed in addition to the desired sequence. For example, the canonical T7 promoter can be referred to by the sequence TAATACGACTCACTATAGG, in which an IVT reaction using the DNA template TAATACGACTCACTATAGGN for the production of desired sequence N will result in the mRNA sequence GG-N. In general, and without wishing to be bound by theory, T7 polymerase more efficiently transcribes RNA transcripts beginning with guanosine. In instances where additional 5' nucleotides are not desired (*e.g.*, no additional GG), the RNA polymerase promoter contained in the DNA template can be a sequence the results in transcripts containing only the 5' nucleotides of the desired sequence, *e.g.*, a SAM having the native 5' sequence of the self-replicating virus from which the SAM vector is derived. For example, a minimal T7 promoter can be referred to by the sequence TAATACGACTCACTATA, in which an IVT reaction using the DNA template TAATACGACTCACTATAN for the production of desired sequence N will result in the mRNA sequence N. Likewise, a minimal SP6 promoter referred to by the sequence ATTTAGGTGACACTATA can be used to generate transcripts without additional 5' nucleotides. In a typical IVT reaction, the DNA template is incubated with the appropriate RNA polymerase enzyme, buffer agents, and nucleotides (NTPs).

[00311] The resulting RNA polynucleotide can optionally be further modified including, but limited to, addition of a 5' cap structure such as 7-methylguanosine or a related structure, and optionally modifying the 3' end to include a polyadenylate (polyA) tail. In a modified IVT reaction, RNA is capped with a 5' cap structure co-transcriptionally through the addition of cap analogues during IVT. Cap analogues can include dinucleotide (m⁷G-ppp-N) cap analogues or

trinucleotide (m⁷G-ppp-N-N) cap analogues, where N represents a nucleotide or modified nucleotide (e.g., ribonucleosides including, but not limited to, adenosine, guanosine, cytidine, and uridine). Exemplary cap analogues and their use in IVT reactions are also described in U.S. Pat. No. 10,519,189, herein incorporated by reference for all purposes. As discussed, T7 polymerase more efficiently transcribes RNA transcripts beginning with guanosine. To improve transcription efficiency in templates that do not begin with guanosine, a trinucleotide cap analogue (m⁷G-ppp-N-N) can be used. The trinucleotide cap analogue can increase transcription efficiency 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20-fold or more relative to an IVT reaction using a dinucleotide cap analogue (m⁷G-ppp-N).

[00312] A 5' cap structure can also be added following transcription, such as using a vaccinia capping system (e.g., NEB Cat. No. M2080) containing mRNA 2'-O-methyltransferase and S-Adenosyl methionine.

[00313] The resulting RNA polynucleotide can optionally be further modified separately from or in addition to the capping techniques described including, but limited to, modifying the 3' end to include a polyadenylate (polyA) tail.

[00314] The RNA can then be purified using techniques well-known in the field, such as phenol-chloroform extraction.

IV.D.4. Delivery via lipid nanoparticle

[00315] An important aspect to consider in vaccine vector design is immunity against the vector itself (Riley 2017). This may be in the form of preexisting immunity to the vector itself, such as with certain human adenovirus systems, or in the form of developing immunity to the vector following administration of the vaccine. The latter is an important consideration if multiple administrations of the same vaccine are performed, such as separate priming and boosting doses, or if the same vaccine vector system is to be used to deliver different antigen cassettes.

[00316] In the case of alphavirus vectors, the standard delivery method is the previously discussed helper virus system that provides capsid, E1, and E2 proteins *in trans* to produce infectious viral particles. However, it is important to note that the E1 and E2 proteins are often major targets of neutralizing antibodies (Strauss 1994). Thus, the efficacy of using alphavirus vectors to deliver antigens of interest to target cells may be reduced if infectious particles are targeted by neutralizing antibodies.

[00317] An alternative to viral particle mediated gene delivery is the use of nanomaterials to deliver expression vectors (Riley 2017). Nanomaterial vehicles, importantly, can be made of

non-immunogenic materials and generally avoid eliciting immunity to the delivery vector itself. These materials can include, but are not limited to, lipids, inorganic nanomaterials, and other polymeric materials. Lipids can be cationic, anionic, or neutral. The materials can be synthetic or naturally derived, and in some instances biodegradable. Lipids can include fats, cholesterol, phospholipids, lipid conjugates including, but not limited to, polyethyleneglycol (PEG) conjugates (PEGylated lipids), waxes, oils, glycerides, and fat soluble vitamins.

[00318] Lipid nanoparticles (LNPs) are an attractive delivery system due to the amphiphilic nature of lipids enabling formation of membranes and vesicle like structures (Riley 2017). In general, these vesicles deliver the expression vector by absorbing into the membrane of target cells and releasing nucleic acid into the cytosol. In addition, LNPs can be further modified or functionalized to facilitate targeting of specific cell types. Another consideration in LNP design is the balance between targeting efficiency and cytotoxicity. Lipid compositions generally include defined mixtures of cationic, neutral, anionic, and amphipathic lipids. In some instances, specific lipids are included to prevent LNP aggregation, prevent lipid oxidation, or provide functional chemical groups that facilitate attachment of additional moieties. Lipid composition can influence overall LNP size and stability. In an example, the lipid composition comprises dilinoleylmethyl- 4-dimethylaminobutyrate (MC3) or MC3-like molecules. MC3 and MC3-like lipid compositions can be formulated to include one or more other lipids, such as a PEG or PEG-conjugated lipid, a sterol, or neutral lipids.

[00319] Nucleic-acid vectors, such as expression vectors, exposed directly to serum can have several undesirable consequences, including degradation of the nucleic acid by serum nucleases or off-target stimulation of the immune system by the free nucleic acids. Therefore, encapsulation of the alphavirus vector can be used to avoid degradation, while also avoiding potential off-target effects. In certain examples, an alphavirus vector is fully encapsulated within the delivery vehicle, such as within the aqueous interior of an LNP. Encapsulation of the alphavirus vector within an LNP can be carried out by techniques well-known to those skilled in the art, such as microfluidic mixing and droplet generation carried out on a microfluidic droplet generating device. Such devices include, but are not limited to, standard T-junction devices or flow-focusing devices. In an example, the desired lipid formulation, such as MC3 or MC3-like containing compositions, is provided to the droplet generating device in parallel with the alphavirus delivery vector and other desired agents, such that the delivery vector and desired agents are fully encapsulated within the interior of the MC3 or MC3-like based LNP. In an example, the droplet generating device can control the size range and size distribution of the

LNPs produced. For example, the LNP can have a size ranging from 1 to 1000 nanometers in diameter, e.g., 1, 10, 50, 100, 500, or 1000 nanometers. Following droplet generation, the delivery vehicles encapsulating the expression vectors can be further treated or modified to prepare them for administration.

IV.E. Chimpanzee adenovirus (ChAd)

IV.E.1. Viral delivery with chimpanzee adenovirus

[00320] Vaccine compositions for delivery of one or more antigens (e.g., via an antigen cassette and including pathogen-derived peptides, a virally-derived peptides, a bacterially-derived peptides, a fungally-derived peptides, and a parasitically-derived peptides) can be created by providing adenovirus nucleotide sequences of chimpanzee origin, a variety of novel vectors, and cell lines expressing chimpanzee adenovirus genes. A nucleotide sequence of a chimpanzee C68 adenovirus (also referred to herein as ChAdV68) can be used in a vaccine composition for antigen delivery (*See* SEQ ID NO: 1). Use of C68 adenovirus derived vectors is described in further detail in USPN 6,083,716, which is herein incorporated by reference in its entirety, for all purposes.

[00321] In a further aspect, provided herein is a recombinant adenovirus comprising the DNA sequence of a chimpanzee adenovirus such as C68 and an antigen cassette operatively linked to regulatory sequences directing its expression. The recombinant virus is capable of infecting a mammalian, preferably a human, cell and capable of expressing the antigen cassette product in the cell. In this vector, the native chimpanzee E1 gene, and/or E3 gene, and/or E4 gene can be deleted. an antigen cassette can be inserted into any of these sites of gene deletion. The antigen cassette can include an antigen against which a primed immune response is desired.

[00322] In another aspect, provided herein is a mammalian cell infected with a chimpanzee adenovirus such as C68.

[00323] In still a further aspect, a novel mammalian cell line is provided which expresses a chimpanzee adenovirus gene (e.g., from C68) or functional fragment thereof.

[00324] In still a further aspect, provided herein is a method for delivering an antigen cassette into a mammalian cell comprising the step of introducing into the cell an effective amount of a chimpanzee adenovirus, such as C68, that has been engineered to express the antigen cassette.

[00325] Still another aspect provides a method for eliciting an immune response in a mammalian host. The method can comprise the step of administering to the host an effective amount of a recombinant chimpanzee adenovirus, such as C68, comprising an antigen cassette that encodes one or more antigens from the infection against which the immune response is targeted.

[00326] Still another aspect provides a method for eliciting an immune response in a mammalian host to treat or prevent a disease in a subject, such as an infectious disease. The method can comprise the step of administering to the host an effective amount of a recombinant chimpanzee adenovirus, such as C68, comprising an antigen cassette that encodes one or more antigens, such as from the infectious disease against which the immune response is targeted.

[00327] Also disclosed is a non-simian mammalian cell that expresses a chimpanzee adenovirus gene obtained from the sequence of SEQ ID NO: 1. The gene can be selected from the group consisting of the adenovirus E1A, E1B, E2A, E2B, E3, E4, L1, L2, L3, L4 and L5 of SEQ ID NO: 1.

[00328] Also disclosed is a nucleic acid molecule comprising a chimpanzee adenovirus DNA sequence comprising a gene obtained from the sequence of SEQ ID NO: 1. The gene can be selected from the group consisting of said chimpanzee adenovirus E1A, E1B, E2A, E2B, E3, E4, L1, L2, L3, L4 and L5 genes of SEQ ID NO: 1. In some aspects the nucleic acid molecule comprises SEQ ID NO: 1. In some aspects the nucleic acid molecule comprises the sequence of SEQ ID NO: 1, lacking at least one gene selected from the group consisting of E1A, E1B, E2A, E2B, E3, E4, L1, L2, L3, L4 and L5 genes of SEQ ID NO: 1.

[00329] Also disclosed is a vector comprising a chimpanzee adenovirus DNA sequence obtained from SEQ ID NO: 1 and an antigen cassette operatively linked to one or more regulatory sequences which direct expression of the cassette in a heterologous host cell, optionally wherein the chimpanzee adenovirus DNA sequence comprises at least the *cis*-elements necessary for replication and virion encapsidation, the *cis*-elements flanking the antigen cassette and regulatory sequences. In some aspects, the chimpanzee adenovirus DNA sequence comprises a gene selected from the group consisting of E1A, E1B, E2A, E2B, E3, E4, L1, L2, L3, L4 and L5 gene sequences of SEQ ID NO: 1. In some aspects the vector can lack the E1A and/or E1B gene.

[00330] Also disclosed herein is an adenovirus vector comprising: a partially deleted E4 gene comprising a deleted or partially-deleted E4orf2 region and a deleted or partially-deleted E4orf3 region, and optionally a deleted or partially-deleted E4orf4 region. The partially deleted E4 can comprise an E4 deletion of at least nucleotides 34,916 to 35,642 of the sequence shown in SEQ ID NO:1, and wherein the vector comprises at least nucleotides 2 to 36,518 of the sequence set forth in SEQ ID NO:1. The partially deleted E4 can comprise an E4 deletion of at least a partial deletion of nucleotides 34,916 to 34,942 of the sequence shown in SEQ ID NO:1, at least a partial deletion of nucleotides 34,952 to 35,305 of the sequence shown in SEQ ID NO:1, and at least a partial deletion of nucleotides 35,302 to 35,642 of the sequence shown in SEQ ID NO:1, and wherein the vector comprises at least nucleotides 2 to 36,518 of the sequence set forth in SEQ ID NO:1. The partially deleted E4 can comprise an E4 deletion of at least nucleotides 34,980 to 36,516 of the sequence shown in SEQ ID NO:1, and wherein the vector comprises at least nucleotides 2 to 36,518 of the sequence set forth in SEQ ID NO:1. The partially deleted E4 can comprise an E4 deletion of at least nucleotides 34,979 to 35,642 of the sequence shown in SEQ ID NO:1, and wherein the vector comprises at least nucleotides 2 to 36,518 of the sequence set forth in SEQ ID NO:1. The partially deleted E4 can comprise an E4 deletion of at least a partial deletion of E4Orf2, a fully deleted E4Orf3, and at least a partial deletion of E4Orf4. The partially deleted E4 can comprise an E4 deletion of at least a partial deletion of E4Orf2, at least a partial deletion of E4Orf3, and at least a partial deletion of E4Orf4. The partially deleted E4 can comprise an E4 deletion of at least a partial deletion of E4Orf1, a fully deleted E4Orf2, and at least a partial deletion of E4Orf3. The partially deleted E4 can comprise an E4 deletion of at least a partial deletion of E4Orf2 and at least a partial deletion of E4Orf3. The partially deleted E4 can comprise an E4 deletion between the start site of E4Orf1 to the start site of E4Orf5. The partially deleted E4 can be an E4 deletion adjacent to the start site of E4Orf1. The partially deleted E4 can be an E4 deletion adjacent to the start site of E4Orf2. The partially deleted E4 can be an E4 deletion adjacent to the start site of E4Orf3. The partially deleted E4 can be an E4 deletion adjacent to the start site of E4Orf4. The E4 deletion can be at least 50, at least 100, at least 200, at least 300, at least 400, at least 500, at least 600, at least 700, at least 800, at least 900, at least 1000, at least 1100, at least 1200, at least 1300, at least 1400, at least 1500, at least 1600, at least 1700, at least 1800, at least 1900, or at least 2000 nucleotides. The E4 deletion can be at least 700 nucleotides. The E4 deletion can be at least 1500 nucleotides. The E4 deletion can be 50 or less, 100 or less, 200 or less, 300 or less, 400 or less, 500 or less, 600 or less, 700 or less, 800 or less, 900 or less, 1000 or less,

1100 or less, 1200 or less, 1300 or less, 1400 or less, 1500 or less, 1600 or less, 1700 or less, 1800 or less, 1900 or less, or 2000 or less nucleotides. The E4 deletion can be 750 nucleotides or less. The E4 deletion can be at least 1550 nucleotides or less.

[00331] The partially deleted E4 gene can be the E4 gene sequence shown in SEQ ID NO:1 that lacks at least nucleotides 34,916 to 35,642 of the sequence shown in SEQ ID NO:1. The partially deleted E4 gene can be the E4 gene sequence shown in SEQ ID NO:1 that lacks the E4 gene sequence shown in SEQ ID NO:1 and that lacks at least nucleotides 34,916 to 34,942, nucleotides 34,952 to 35,305 of the sequence shown in SEQ ID NO:1, and nucleotides 35,302 to 35,642 of the sequence shown in SEQ ID NO:1. The partially deleted E4 gene can be the E4 gene sequence shown in SEQ ID NO:1 and that lacks at least nucleotides 34,980 to 36,516 of the sequence shown in SEQ ID NO:1. The partially deleted E4 gene can be the E4 gene sequence shown in SEQ ID NO:1 and that lacks at least nucleotides 34,979 to 35,642 of the sequence shown in SEQ ID NO:1. The adenovirus vector having the partially deleted E4 gene can have a cassette, wherein the cassette comprises at least one payload nucleic acid sequence, and wherein the cassette comprises at least one promoter sequence operably linked to the at least one payload nucleic acid sequence. The adenovirus vector having the partially deleted E4 gene can have one or more genes or regulatory sequences of the ChAdV68 sequence shown in SEQ ID NO: 1, optionally wherein the one or more genes or regulatory sequences comprise at least one of the chimpanzee adenovirus inverted terminal repeat (ITR), E1A, E1B, E2A, E2B, E3, E4, L1, L2, L3, L4, and L5 genes of the sequence shown in SEQ ID NO: 1. The adenovirus vector having the partially deleted E4 gene can have nucleotides 2 to 34,916 of the sequence shown in SEQ ID NO:1, wherein the partially deleted E4 gene is 3' of the nucleotides 2 to 34,916, and optionally the nucleotides 2 to 34,916 additionally lack nucleotides 577 to 3403 of the sequence shown in SEQ ID NO:1 corresponding to an E1 deletion and/or lack nucleotides 27,125 to 31,825 of the sequence shown in SEQ ID NO:1 corresponding to an E3 deletion. The adenovirus vector having the partially deleted E4 gene can have nucleotides 35,643 to 36,518 of the sequence shown in SEQ ID NO:1, and wherein the partially deleted E4 gene is 5' of the nucleotides 35,643 to 36,518. The adenovirus vector having the partially deleted E4 gene can have nucleotides 2 to 34,916 of the sequence shown in SEQ ID NO:1, wherein the partially deleted E4 gene is 3' of the nucleotides 2 to 34,916, the nucleotides 2 to 34,916 additionally lack nucleotides 577 to 3403 of the sequence shown in SEQ ID NO:1 corresponding to an E1 deletion and lack nucleotides 27,125 to 31,825 of the sequence shown in SEQ ID NO:1 corresponding to an E3 deletion. The adenovirus vector having the partially deleted E4 gene

can have nucleotides 2 to 34,916 of the sequence shown in SEQ ID NO:1, wherein the partially deleted E4 gene is 3' of the nucleotides 2 to 34,916, the nucleotides 2 to 34,916 additionally lack nucleotides 577 to 3403 of the sequence shown in SEQ ID NO:1 corresponding to an E1 deletion and lack nucleotides 27,125 to 31,825 of the sequence shown in SEQ ID NO:1 corresponding to an E3 deletion, and have nucleotides 35,643 to 36,518 of the sequence shown in SEQ ID NO:1, and wherein the partially deleted E4 gene is 5' of the nucleotides 35,643 to 36,518.

[00332] The partially deleted E4 gene can be the E4 gene sequence shown in SEQ ID NO:1 that lacks at least nucleotides 34,916 to 35,642 of the sequence shown in SEQ ID NO:1, nucleotides 2 to 34,916 of the sequence shown in SEQ ID NO:1, wherein the partially deleted E4 gene is 3' of the nucleotides 2 to 34,916, the nucleotides 2 to 34,916 additionally lack nucleotides 577 to 3403 of the sequence shown in SEQ ID NO:1 corresponding to an E1 deletion and lack nucleotides 27,125 to 31,825 of the sequence shown in SEQ ID NO:1 corresponding to an E3 deletion, and have nucleotides 35,643 to 36,518 of the sequence shown in SEQ ID NO:1, and wherein the partially deleted E4 gene is 5' of the nucleotides 35,643 to 36,518.

[00333] Also disclosed herein is a host cell transfected with a vector disclosed herein such as a C68 vector engineered to expression an antigen cassette. Also disclosed herein is a human cell that expresses a selected gene introduced therein through introduction of a vector disclosed herein into the cell.

[00334] Also disclosed herein is a method for delivering an antigen cassette to a mammalian cell comprising introducing into said cell an effective amount of a vector disclosed herein such as a C68 vector engineered to expression the antigen cassette.

[00335] Also disclosed herein is a method for producing an antigen comprising introducing a vector disclosed herein into a mammalian cell, culturing the cell under suitable conditions and producing the antigen.

IV.E.2. E1-Expressing Complementation Cell Lines

[00336] To generate recombinant chimpanzee adenoviruses (Ad) deleted in any of the genes described herein, the function of the deleted gene region, if essential to the replication and infectivity of the virus, can be supplied to the recombinant virus by a helper virus or cell line, i.e., a complementation or packaging cell line. For example, to generate a replication-defective chimpanzee adenovirus vector, a cell line can be used which expresses the E1 gene products of the human or chimpanzee adenovirus; such a cell line can include HEK293 or variants thereof. The protocol for the generation of the cell lines expressing the chimpanzee E1 gene products (Examples 3 and 4 of USPN 6,083,716) can be followed to generate a cell line which expresses any selected chimpanzee adenovirus gene.

[00337] An AAV augmentation assay can be used to identify a chimpanzee adenovirus E1-expressing cell line. This assay is useful to identify E1 function in cell lines made by using the E1 genes of other uncharacterized adenoviruses, e.g., from other species. That assay is described in Example 4B of USPN 6,083,716.

[00338] A selected chimpanzee adenovirus gene, e.g., E1, can be under the transcriptional control of a promoter for expression in a selected parent cell line. Inducible or constitutive promoters can be employed for this purpose. Among inducible promoters are included the sheep metallothionine promoter, inducible by zinc, or the mouse mammary tumor virus (MMTV) promoter, inducible by a glucocorticoid, particularly, dexamethasone. Other inducible promoters, such as those identified in International patent application WO95/13392, incorporated by reference herein can also be used in the production of packaging cell lines. Constitutive promoters in control of the expression of the chimpanzee adenovirus gene can be employed also.

[00339] A parent cell can be selected for the generation of a novel cell line expressing any desired C68 gene. Without limitation, such a parent cell line can be HeLa [ATCC Accession No. CCL 2], A549 [ATCC Accession No. CCL 185], KB [CCL 17], Detroit [e.g., Detroit 510, CCL 72] and WI-38 [CCL 75] cells. Other suitable parent cell lines can be obtained from other sources. Parent cell lines can include CHO, HEK293 or variants thereof, 911, HeLa, A549, LP-293, PER.C6, or AE1-2a.

[00340] An E1-expressing cell line can be useful in the generation of recombinant chimpanzee adenovirus E1 deleted vectors. Cell lines constructed using essentially the same procedures that express one or more other chimpanzee adenoviral gene products are useful in

the generation of recombinant chimpanzee adenovirus vectors deleted in the genes that encode those products. Further, cell lines which express other human Ad E1 gene products are also useful in generating chimpanzee recombinant Ads.

IV.E.3. Recombinant Viral Particles as Vectors

[00341] The compositions disclosed herein can comprise viral vectors, that deliver at least one antigen to cells. Such vectors comprise a chimpanzee adenovirus DNA sequence such as C68 and an antigen cassette operatively linked to regulatory sequences which direct expression of the cassette. The C68 vector is capable of expressing the cassette in an infected mammalian cell. The C68 vector can be functionally deleted in one or more viral genes. an antigen cassette comprises at least one antigen under the control of one or more regulatory sequences such as a promoter. Optional helper viruses and/or packaging cell lines can supply to the chimpanzee viral vector any necessary products of deleted adenoviral genes.

[00342] The term "functionally deleted" means that a sufficient amount of the gene region is removed or otherwise altered, e.g., by mutation or modification, so that the gene region is no longer capable of producing one or more functional products of gene expression. Mutations or modifications that can result in functional deletions include, but are not limited to, nonsense mutations such as introduction of premature stop codons and removal of canonical and non-canonical start codons, mutations that alter mRNA splicing or other transcriptional processing, or combinations thereof. If desired, the entire gene region can be removed.

[00343] Modifications of the nucleic acid sequences forming the vectors disclosed herein, including sequence deletions, insertions, and other mutations may be generated using standard molecular biological techniques and are within the scope of this invention.

IV.E.4. Construction of The Viral Plasmid Vector

[00344] The chimpanzee adenovirus C68 vectors useful in this invention include recombinant, defective adenoviruses, that is, chimpanzee adenovirus sequences functionally deleted in the E1a or E1b genes, and optionally bearing other mutations, e.g., temperature-sensitive mutations or deletions in other genes. It is anticipated that these chimpanzee sequences are also useful in forming hybrid vectors from other adenovirus and/or adeno-associated virus sequences. Homologous adenovirus vectors prepared from human adenoviruses are described in the published literature [see, for example, Kozarsky I and II, cited above, and references cited therein, U.S. Pat. No. 5,240,846].

[00345] In the construction of useful chimpanzee adenovirus C68 vectors for delivery of an antigen cassette to a human (or other mammalian) cell, a range of adenovirus nucleic acid sequences can be employed in the vectors. A vector comprising minimal chimpanzee C68 adenovirus sequences can be used in conjunction with a helper virus to produce an infectious recombinant virus particle. The helper virus provides essential gene products required for viral infectivity and propagation of the minimal chimpanzee adenoviral vector. When only one or more selected deletions of chimpanzee adenovirus genes are made in an otherwise functional viral vector, the deleted gene products can be supplied in the viral vector production process by propagating the virus in a selected packaging cell line that provides the deleted gene functions in trans.

IV.E.5. Recombinant Minimal Adenovirus

[00346] A minimal chimpanzee Ad C68 virus is a viral particle containing just the adenovirus cis-elements necessary for replication and virion encapsidation. That is, the vector contains the cis-acting 5' and 3' inverted terminal repeat (ITR) sequences of the adenoviruses (which function as origins of replication) and the native 5' packaging/enhancer domains (that contain sequences necessary for packaging linear Ad genomes and enhancer elements for the E1 promoter). See, for example, the techniques described for preparation of a "minimal" human Ad vector in International Patent Application WO96/13597 and incorporated herein by reference.

IV.E.6. Other Defective Adenoviruses

[00347] Recombinant, replication-deficient adenoviruses can also contain more than the minimal chimpanzee adenovirus sequences. These other Ad vectors can be characterized by deletions of various portions of gene regions of the virus, and infectious virus particles formed by the optional use of helper viruses and/or packaging cell lines.

[00348] As one example, suitable vectors may be formed by deleting all or a sufficient portion of the C68 adenoviral immediate early gene E1a and delayed early gene E1b, so as to eliminate their normal biological functions. Replication-defective E1-deleted viruses are capable of replicating and producing infectious virus when grown on a chimpanzee adenovirus-transformed, complementation cell line containing functional adenovirus E1a and E1b genes which provide the corresponding gene products in trans. Based on the homologies to known adenovirus sequences, it is anticipated that, as is true for the human recombinant E1-deleted adenoviruses of the art, the resulting recombinant chimpanzee adenovirus is capable of

infecting many cell types and can express antigen(s), but cannot replicate in most cells that do not carry the chimpanzee E1 region DNA unless the cell is infected at a very high multiplicity of infection.

[00349] As another example, all or a portion of the C68 adenovirus delayed early gene E3 can be eliminated from the chimpanzee adenovirus sequence which forms a part of the recombinant virus.

[00350] Chimpanzee adenovirus C68 vectors can also be constructed having a deletion of the E4 gene. Still another vector can contain a deletion in the delayed early gene E2a.

[00351] Deletions can also be made in any of the late genes L1 through L5 of the chimpanzee C68 adenovirus genome. Similarly, deletions in the intermediate genes IX and IVa2 can be useful for some purposes. Other deletions may be made in the other structural or non-structural adenovirus genes.

[00352] The above discussed deletions can be used individually, i.e., an adenovirus sequence can contain deletions of E1 only. Alternatively, deletions of entire genes or portions thereof effective to destroy or reduce their biological activity can be used in any combination. For example, in one exemplary vector, the adenovirus C68 sequence can have deletions of the E1 genes and the E4 gene, or of the E1, E2a and E3 genes, or of the E1 and E3 genes, or of E1, E2a and E4 genes, with or without deletion of E3, and so on. As discussed above, such deletions can be used in combination with other mutations, such as temperature-sensitive mutations, to achieve a desired result.

[00353] The cassette comprising antigen(s) be inserted optionally into any deleted region of the chimpanzee C68 Ad virus. Alternatively, the cassette can be inserted into an existing gene region to disrupt the function of that region, if desired.

IV.E.7. Helper Viruses

[00354] Depending upon the chimpanzee adenovirus gene content of the viral vectors employed to carry the antigen cassette, a helper adenovirus or non-replicating virus fragment can be used to provide sufficient chimpanzee adenovirus gene sequences to produce an infective recombinant viral particle containing the cassette.

[00355] Useful helper viruses contain selected adenovirus gene sequences not present in the adenovirus vector construct and/or not expressed by the packaging cell line in which the vector is transfected. A helper virus can be replication-defective and contain a variety of adenovirus

genes in addition to the sequences described above. The helper virus can be used in combination with the E1-expressing cell lines described herein.

[00356] For C68, the "helper" virus can be a fragment formed by clipping the C terminal end of the C68 genome with SspI, which removes about 1300 bp from the left end of the virus. This clipped virus is then co-transfected into an E1-expressing cell line with the plasmid DNA, thereby forming the recombinant virus by homologous recombination with the C68 sequences in the plasmid.

[00357] Helper viruses can also be formed into poly-cation conjugates as described in Wu et al, J. Biol. Chem., 264:16985-16987 (1989); K. J. Fisher and J. M. Wilson, Biochem. J., 299:49 (Apr. 1, 1994). Helper virus can optionally contain a reporter gene. A number of such reporter genes are known to the art. The presence of a reporter gene on the helper virus which is different from the antigen cassette on the adenovirus vector allows both the Ad vector and the helper virus to be independently monitored. This second reporter is used to enable separation between the resulting recombinant virus and the helper virus upon purification.

IV.E.8. Assembly of Viral Particle and Infection of a Cell Line

[00358] Assembly of the selected DNA sequences of the adenovirus, the antigen cassette, and other vector elements into various intermediate plasmids and shuttle vectors, and the use of the plasmids and vectors to produce a recombinant viral particle can all be achieved using conventional techniques. Such techniques include conventional cloning techniques of cDNA, in vitro recombination techniques (e.g., Gibson assembly), use of overlapping oligonucleotide sequences of the adenovirus genomes, polymerase chain reaction, and any suitable method which provides the desired nucleotide sequence. Standard transfection and co-transfection techniques are employed, e.g., CaPO₄ precipitation techniques or liposome-mediated transfection methods such as lipofectamine. Other conventional methods employed include homologous recombination of the viral genomes, plaquing of viruses in agar overlay, methods of measuring signal generation, and the like.

[00359] For example, following the construction and assembly of the desired antigen cassette-containing viral vector, the vector can be transfected in vitro in the presence of a helper virus into the packaging cell line. Homologous recombination occurs between the helper and the vector sequences, which permits the adenovirus-antigen sequences in the vector to be replicated and packaged into virion capsids, resulting in the recombinant viral vector particles.

[00360] The resulting recombinant chimpanzee C68 adenoviruses are useful in transferring an antigen cassette to a selected cell. In in vivo experiments with the recombinant virus grown in the packaging cell lines, the E1-deleted recombinant chimpanzee adenovirus demonstrates utility in transferring a cassette to a non-chimpanzee, preferably a human, cell.

IV.E.9. Use of the Recombinant Virus Vectors

[00361] The resulting recombinant chimpanzee C68 adenovirus containing the antigen cassette (produced by cooperation of the adenovirus vector and helper virus or adenoviral vector and packaging cell line, as described above) thus provides an efficient gene transfer vehicle which can deliver antigen(s) to a subject in vivo or ex vivo.

[00362] The above-described recombinant vectors are administered to humans according to published methods for gene therapy. A chimpanzee viral vector bearing an antigen cassette can be administered to a patient, preferably suspended in a biologically compatible solution or pharmaceutically acceptable delivery vehicle. A suitable vehicle includes sterile saline. Other aqueous and non-aqueous isotonic sterile injection solutions and aqueous and non-aqueous sterile suspensions known to be pharmaceutically acceptable carriers and well known to those of skill in the art may be employed for this purpose.

[00363] The chimpanzee adenoviral vectors are administered in sufficient amounts to transduce the human cells and to provide sufficient levels of antigen transfer and expression to provide a therapeutic benefit without undue adverse or with medically acceptable physiological effects, which can be determined by those skilled in the medical arts. Conventional and pharmaceutically acceptable routes of administration include, but are not limited to, direct delivery to the liver, intranasal, intravenous, intramuscular, subcutaneous, intradermal, oral and other parental routes of administration. Routes of administration may be combined, if desired.

[00364] Dosages of the viral vector will depend primarily on factors such as the condition being treated, the age, weight and health of the patient, and may thus vary among patients. The dosage will be adjusted to balance the therapeutic benefit against any side effects and such dosages may vary depending upon the therapeutic application for which the recombinant vector is employed. The levels of expression of antigen(s) can be monitored to determine the frequency of dosage administration.

[00365] Recombinant, replication defective adenoviruses can be administered in a "pharmaceutically effective amount", that is, an amount of recombinant adenovirus that is effective in a route of administration to transfect the desired cells and provide sufficient levels

of expression of the selected gene to provide a vaccinal benefit, i.e., some measurable level of protective immunity. C68 vectors comprising an antigen cassette can be co-administered with adjuvant. Adjuvant can be separate from the vector (e.g., alum) or encoded within the vector, in particular if the adjuvant is a protein. Adjuvants are well known in the art.

[00366] Conventional and pharmaceutically acceptable routes of administration include, but are not limited to, intranasal, intramuscular, intratracheal, subcutaneous, intradermal, rectal, oral and other parental routes of administration. Routes of administration may be combined, if desired, or adjusted depending upon the immunogen or the disease. For example, in prophylaxis of rabies, the subcutaneous, intratracheal and intranasal routes are preferred. The route of administration primarily will depend on the nature of the disease being treated.

[00367] The levels of immunity to antigen(s) can be monitored to determine the need, if any, for boosters. Following an assessment of antibody titers in the serum, for example, optional booster immunizations may be desired

V. Therapeutic and Manufacturing Methods

[00368] Also provided is a method of inducing a infectious disease organism-specific immune response in a subject, vaccinating against a infectious disease organism, treating and or alleviating a symptom of an infection associated with an infectious disease organism in a subject by administering to the subject one or more antigens such as a plurality of antigens identified using methods disclosed herein.

[00369] In some aspects, a subject has been diagnosed with an infection or is at risk of an infection, such as age, geographical/travel, and/or work-related increased risk of or predisposition to an infection, or at risk to a seasonal and/or novel disease infection.

[00370] A antigen can be administered in an amount sufficient to induce a CTL response. an antigen can be administered in an amount sufficient to induce a T cell response response. an antigen can be administered in an amount sufficient to induce a B cell response response.

[00371] A antigen can be administered alone or in combination with other therapeutic agents. Therapeutic agents can include those that target an infectious disease organism, such as an anti-viral or antibiotic agent.

[00372] In addition, a subject can be further administered an anti-immunosuppressive/immunostimulatory agent such as a checkpoint inhibitor. For example, the subject can be further administered an anti-CTLA antibody or anti-PD-1 or anti-PD-L1. Blockade of CTLA-4 or PD-L1 by antibodies can enhance the immune response to cancerous

cells in the patient. In particular, CTLA-4 blockade has been shown effective when following a vaccination protocol.

[00373] The optimum amount of each antigen to be included in a vaccine composition and the optimum dosing regimen can be determined. For example, an antigen or its variant can be prepared for intravenous (i.v.) injection, sub-cutaneous (s.c.) injection, intradermal (i.d.) injection, intraperitoneal (i.p.) injection, intramuscular (i.m.) injection. Methods of injection include s.c., i.d., i.p., i.m., and i.v. Methods of DNA or RNA injection include i.d., i.m., s.c., i.p. and i.v. Other methods of administration of the vaccine composition are known to those skilled in the art.

[00374] A vaccine can be compiled so that the selection, number and/or amount of antigens present in the composition is/are tissue, infectious disease, and/or patient-specific. For instance, the exact selection of peptides can be guided by expression patterns of the parent proteins in a given tissue or guided by mutation or disease status of a patient. The selection can be dependent on the the specific infectious disease (*e.g.* the specific SARS-CoV-2 isolate the subject is infected with or at risk for infection by), the status of the disease, the goal of the vaccination (*e.g.*, preventative or targeting an ongoing disease), earlier treatment regimens, the immune status of the patient, and, of course, the HLA-haplotype of the patient. Furthermore, a vaccine can contain individualized components, according to personal needs of the particular patient. Examples include varying the selection of antigens according to the expression of the antigen in the particular patient or adjustments for secondary treatments following a first round or scheme of treatment.

[00375] A patient can be identified for administration of an antigen vaccine through the use of various diagnostic methods, *e.g.*, patient selection methods described further below. Patient selection can involve identifying mutations in, or expression patterns of, one or more genes. Patient selection can involve identifying the infectious disease of an ongoing infection. Patient selection can involve identifying risk of an infection by an infectious disease. In some cases, patient selection involves identifying the haplotype of the patient. The various patient selection methods can be performed in parallel, *e.g.*, a sequencing diagnostic can identify both the mutations and the haplotype of a patient. The various patient selection methods can be performed sequentially, *e.g.*, one diagnostic test identifies the mutations and separate diagnostic test identifies the haplotype of a patient, and where each test can be the same (*e.g.*, both high-throughput sequencing) or different (*e.g.*, one high-throughput sequencing and the other Sanger sequencing) diagnostic methods.

[00376] For a composition to be used as a vaccine for an infectious disease, antigens with similar normal self-peptides that are expressed in high amounts in normal tissues can be avoided or be present in low amounts in a composition described herein. On the other hand, if it is known that the infected cell of a patient expresses high amounts of a certain antigen, the respective pharmaceutical composition for treatment of this infection can be present in high amounts and/or more than one antigen specific for this particular antigen or pathway of this antigen can be included.

[00377] Compositions comprising an antigen can be administered to an individual already suffering from an infection. In therapeutic applications, compositions are administered to a patient in an amount sufficient to elicit an effective CTL response to the infectious disease organism antigen and to cure or at least partially arrest symptoms and/or complications. An amount adequate to accomplish this is defined as "therapeutically effective dose." Amounts effective for this use will depend on, e.g., the composition, the manner of administration, the stage and severity of the disease being treated, the weight and general state of health of the patient, and the judgment of the prescribing physician. It should be kept in mind that compositions can generally be employed in serious disease states, that is, life-threatening or potentially life threatening situations, especially when the infectious disease organism has induced organ damage and/or other immune pathology. In such cases, in view of the minimization of extraneous substances and the relative nontoxic nature of an antigen, it is possible and can be felt desirable by the treating physician to administer substantial excesses of these compositions.

[00378] For therapeutic use, administration can begin at the detection or treatment of an infection. This can be followed by boosting doses until at least symptoms are substantially abated and for a period thereafter or immunity is considered to be provided (e.g., a memory B cell or T cell population, or antigen specific B cells or antibodies are produced).

[00379] The pharmaceutical compositions (e.g., vaccine compositions) for therapeutic treatment are intended for parenteral, topical, nasal, oral or local administration. A pharmaceutical compositions can be administered parenterally, e.g., intravenously, subcutaneously, intradermally, or intramuscularly. The compositions can be administered to target specific infected tissues and/or cells of a subject.. Disclosed herein are compositions for parenteral administration which comprise a solution of the antigen and vaccine compositions are dissolved or suspended in an acceptable carrier, e.g., an aqueous carrier. A variety of aqueous carriers can be used, e.g., water, buffered water, 0.9% saline, 0.3% glycine, hyaluronic

acid and the like. These compositions can be sterilized by conventional, well known sterilization techniques, or can be sterile filtered. The resulting aqueous solutions can be packaged for use as is, or lyophilized, the lyophilized preparation being combined with a sterile solution prior to administration. The compositions may contain pharmaceutically acceptable auxiliary substances as required to approximate physiological conditions, such as pH adjusting and buffering agents, tonicity adjusting agents, wetting agents and the like, for example, sodium acetate, sodium lactate, sodium chloride, potassium chloride, calcium chloride, sorbitan monolaurate, triethanolamine oleate, etc.

[00380] Antigens can also be administered via liposomes, which target them to a particular cells tissue, such as lymphoid tissue. Liposomes are also useful in increasing half-life. Liposomes include emulsions, foams, micelles, insoluble monolayers, liquid crystals, phospholipid dispersions, lamellar layers and the like. In these preparations the antigen to be delivered is incorporated as part of a liposome, alone or in conjunction with a molecule which binds to, e.g., a receptor prevalent among lymphoid cells, such as monoclonal antibodies which bind to the CD45 antigen, or with other therapeutic or immunogenic compositions. Thus, liposomes filled with a desired antigen can be directed to the site of lymphoid cells, where the liposomes then deliver the selected therapeutic/immunogenic compositions. Liposomes can be formed from standard vesicle-forming lipids, which generally include neutral and negatively charged phospholipids and a sterol, such as cholesterol. The selection of lipids is generally guided by consideration of, e.g., liposome size, acid lability and stability of the liposomes in the blood stream. A variety of methods are available for preparing liposomes, as described in, e.g., Szoka et al., *Ann. Rev. Biophys. Bioeng.* 9; 467 (1980), U.S. Pat. Nos. 4,235,871, 4,501,728, 4,501,728, 4,837,028, and 5,019,369.

[00381] For targeting to the immune cells, a ligand to be incorporated into the liposome can include, e.g., antibodies or fragments thereof specific for cell surface determinants of the desired immune system cells. A liposome suspension can be administered intravenously, locally, topically, etc. in a dose which varies according to, inter alia, the manner of administration, the peptide being delivered, and the stage of the disease being treated.

[00382] For therapeutic or immunization purposes, nucleic acids encoding a peptide and optionally one or more of the peptides described herein can also be administered to the patient. A number of methods are conveniently used to deliver the nucleic acids to the patient. For instance, the nucleic acid can be delivered directly, as "naked DNA". This approach is described, for instance, in Wolff et al., *Science* 247: 1465-1468 (1990) as well as U.S. Pat. Nos.

5,580,859 and 5,589,466. The nucleic acids can also be administered using ballistic delivery as described, for instance, in U.S. Pat. No. 5,204,253. Particles comprised solely of DNA can be administered. Alternatively, DNA can be adhered to particles, such as gold particles. Approaches for delivering nucleic acid sequences can include viral vectors, mRNA vectors, and DNA vectors with or without electroporation.

[00383] The nucleic acids can also be delivered complexed to cationic compounds, such as cationic lipids. Lipid-mediated gene delivery methods are described, for instance, in 9618372WOAWO 96/18372; 9324640WOAWO 93/24640; Mannino & Gould-Fogerite, *BioTechniques* 6(7): 682-691 (1988); U.S. Pat. No. 5,279,833 Rose U.S. Pat. No. 5,279,833; 9106309WOAWO 91/06309; and Felgner et al., *Proc. Natl. Acad. Sci. USA* 84: 7413-7414 (1987).

[00384] Antigens can also be included in viral vector-based vaccine platforms, such as vaccinia, fowlpox, self-replicating alphavirus, marabavirus, adenovirus (*See, e.g., Tatsis et al., Adenoviruses, Molecular Therapy* (2004) 10, 616—629), or lentivirus, including but not limited to second, third or hybrid second/third generation lentivirus and recombinant lentivirus of any generation designed to target specific cell types or receptors (*See, e.g., Hu et al., Immunization Delivered by Lentiviral Vectors for Cancer and Infectious Diseases, Immunol Rev.* (2011) 239(1): 45-61, Sakuma et al., *Lentiviral vectors: basic to translational, Biochem J.* (2012) 443(3):603-18, Cooper et al., *Rescue of splicing-mediated intron loss maximizes expression in lentiviral vectors containing the human ubiquitin C promoter, Nucl. Acids Res.* (2015) 43 (1): 682-690, Zufferey et al., *Self-Inactivating Lentivirus Vector for Safe and Efficient In Vivo Gene Delivery, J. Virol.* (1998) 72 (12): 9873-9880). Dependent on the packaging capacity of the above mentioned viral vector-based vaccine platforms, this approach can deliver one or more nucleotide sequences that encode one or more antigen peptides. The sequences may be flanked by non-mutated sequences, may be separated by linkers or may be preceded with one or more sequences targeting a subcellular compartment (*See, e.g., Gros et al., Prospective identification of neoantigen-specific lymphocytes in the peripheral blood of melanoma patients, Nat Med.* (2016) 22 (4):433-8, Stronen et al., *Targeting of cancer neoantigens with donor-derived T cell receptor repertoires, Science.* (2016) 352 (6291):1337-41, Lu et al., *Efficient identification of mutated cancer antigens recognized by T cells associated with durable tumor regressions, Clin Cancer Res.* (2014) 20(13):3401-10). Upon introduction into a host, infected cells express the antigens, and thereby elicit a host immune (e.g., CTL) response against the peptide(s). Vaccinia vectors and methods useful in

immunization protocols are described in, e.g., U.S. Pat. No. 4,722,848. Another vector is BCG (Bacille Calmette Guerin). BCG vectors are described in Stover et al. (Nature 351:456-460 (1991)). A wide variety of other vaccine vectors useful for therapeutic administration or immunization of antigens, e.g., Salmonella typhi vectors, and the like will be apparent to those skilled in the art from the description herein.

[00385] A means of administering nucleic acids uses minigene constructs encoding one or multiple epitopes. To create a DNA sequence encoding the selected CTL epitopes (minigene) for expression in human cells, the amino acid sequences of the epitopes are reverse translated. A human codon usage table is used to guide the codon choice for each amino acid. These epitope-encoding DNA sequences are directly adjoined, creating a continuous polypeptide sequence. To optimize expression and/or immunogenicity, additional elements can be incorporated into the minigene design. Examples of amino acid sequence that could be reverse translated and included in the minigene sequence include: helper T lymphocyte, epitopes, a leader (signal) sequence, and an endoplasmic reticulum retention signal. In addition, MHC presentation of CTL epitopes can be improved by including synthetic (e.g. poly-alanine) or naturally-occurring flanking sequences adjacent to the CTL epitopes. The minigene sequence is converted to DNA by assembling oligonucleotides that encode the plus and minus strands of the minigene. Overlapping oligonucleotides (30-100 bases long) are synthesized, phosphorylated, purified and annealed under appropriate conditions using well known techniques. The ends of the oligonucleotides are joined using T4 DNA ligase. This synthetic minigene, encoding the CTL epitope polypeptide, can then be cloned into a desired expression vector.

[00386] Purified plasmid DNA can be prepared for injection using a variety of formulations. The simplest of these is reconstitution of lyophilized DNA in sterile phosphate-buffer saline (PBS). A variety of methods have been described, and new techniques can become available. As noted above, nucleic acids are conveniently formulated with cationic lipids. In addition, glycolipids, fusogenic liposomes, peptides and compounds referred to collectively as protective, interactive, non-condensing (PINC) could also be complexed to purified plasmid DNA to influence variables such as stability, intramuscular dispersion, or trafficking to specific organs or cell types.

[00387] Also disclosed is a method of manufacturing an infectious disease vaccine, comprising performing the steps of a method disclosed herein; and producing an infectious disease vaccine comprising a plurality of antigens or a subset of the plurality of antigens.

[00388] Antigens disclosed herein can be manufactured using methods known in the art. For example, a method of producing an antigen or a vector (e.g., a vector including at least one sequence encoding one or more antigens) disclosed herein can include culturing a host cell under conditions suitable for expressing the antigen or vector wherein the host cell comprises at least one polynucleotide encoding the antigen or vector, and purifying the antigen or vector. Standard purification methods include chromatographic techniques, electrophoretic, immunological, precipitation, dialysis, filtration, concentration, and chromatofocusing techniques.

[00389] Host cells can include a Chinese Hamster Ovary (CHO) cell, NS0 cell, yeast, or a HEK293 cell. Host cells can be transformed with one or more polynucleotides comprising at least one nucleic acid sequence that encodes an antigen or vector disclosed herein, optionally wherein the isolated polynucleotide further comprises a promoter sequence operably linked to the at least one nucleic acid sequence that encodes the antigen or vector. In certain embodiments the isolated polynucleotide can be cDNA.

VI. Antigen Use and Administration

[00390] A vaccination protocol can be used to dose a subject with one or more antigens. A priming vaccine and a boosting vaccine can be used to dose the subject. The priming vaccine can be based on C68 (e.g., the sequences shown in SEQ ID NO:1 or 2) or srRNA (e.g., the sequences shown in SEQ ID NO:3 or 4) and the boosting vaccine can be based on C68 (e.g., the sequences shown in SEQ ID NO:1 or 2) or srRNA (e.g., the sequences shown in SEQ ID NO:3 or 4). Each vector typically includes a cassette that includes antigens. Cassettes can include about 20 antigens, separated by spacers such as the natural sequence that normally surrounds each antigen or other non-natural spacer sequences such as AAY. Cassettes can also include MHCII antigens such as a tetanus toxoid antigen and PADRE antigen, which can be considered universal class II antigens. Cassettes can also include a targeting sequence such as a ubiquitin targeting sequence. In addition, each vaccine dose can be administered to the subject in conjunction with (e.g., concurrently, before, or after) a checkpoint inhibitor (CPI). CPI's can include those that inhibit CTLA4, PD1, and/or PDL1 such as antibodies or antigen-binding portions thereof. Such antibodies can include tremelimumab or durvalumab.

[00391] A priming vaccine can be injected (e.g., intramuscularly) in a subject. Bilateral injections per dose can be used. For example, one or more injections of ChAdV68 (C68) can be used (e.g., total dose 1×10^{12} viral particles); one or more injections of self-replicating RNA

(srRNA) at low vaccine dose selected from the range 0.001 to 1 ug RNA, in particular 0.1 or 1 ug can be used; or one or more injections of srRNA at high vaccine dose selected from the range 1 to 100 ug RNA, in particular 10 or 100 ug can be used.

[00392] A vaccine boost (boosting vaccine) can be injected (e.g., intramuscularly) after prime vaccination. A boosting vaccine can be administered about every 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 weeks, e.g., every 4 weeks and/or 8 weeks after the prime. Bilateral injections per dose can be used. For example, one or more injections of ChAdV68 (C68) can be used (e.g., total dose 1×10^{12} viral particles); one or more injections of self-replicating RNA (srRNA) at low vaccine dose selected from the range 0.001 to 1 ug RNA, in particular 0.1 or 1 ug can be used; or one or more injections of srRNA at high vaccine dose selected from the range 1 to 100 ug RNA, in particular 10 or 100 ug can be used.

[00393] Anti-CTLA-4 (e.g., tremelimumab) can also be administered to the subject. For example, anti-CTLA4 can be administered subcutaneously near the site of the intramuscular vaccine injection (ChAdV68 prime or srRNA low doses) to ensure drainage into the same lymph node. Tremelimumab is a selective human IgG2 mAb inhibitor of CTLA-4. Target Anti-CTLA-4 (tremelimumab) subcutaneous dose is typically 70-75 mg (in particular 75 mg) with a dose range of, e.g., 1-100 mg or 5-420 mg.

[00394] In certain instances an anti-PD-L1 antibody can be used such as durvalumab (MEDI 4736). Durvalumab is a selective, high affinity human IgG1 mAb that blocks PD-L1 binding to PD-1 and CD80. Durvalumab is generally administered at 20 mg/kg i.v. every 4 weeks.

[00395] Immune monitoring can be performed before, during, and/or after vaccine administration. Such monitoring can inform safety and efficacy, among other parameters.

[00396] To perform immune monitoring, PBMCs are commonly used. PBMCs can be isolated before prime vaccination, and after prime vaccination (e.g. 4 weeks and 8 weeks). PBMCs can be harvested just prior to boost vaccinations and after each boost vaccination (e.g. 4 weeks and 8 weeks).

[00397] Immune response, such as T cell responses and B cell responses, can be assessed as part of an immune monitoring protocol. For example, the ability of a vaccine composition described herein to stimulate an immune response can be monitored and/or assessed. As used herein, “stimulate an immune response” refers to any increase in a immune response, such as initiating an immune response (e.g., a priming vaccine stimulating the initiation of an immune response in a naïve subject) or enhancement of an immune response (e.g., a boosting vaccine

stimulating the enhancement of an immune response in a subject having a pre-existing immune response to an antigen, such as a pre-existing immune response initiated by a priming vaccine).

[00398] T cell responses can be measured using one or more methods known in the art such as ELISpot, intracellular cytokine staining, cytokine secretion and cell surface capture, T cell proliferation, MHC multimer staining, or by cytotoxicity assay. T cell responses to epitopes encoded in vaccines can be monitored from PBMCs by measuring induction of cytokines, such as IFN-gamma, using an ELISpot assay. Specific CD4 or CD8 T cell responses to epitopes encoded in vaccines can be monitored from PBMCs by measuring induction of cytokines captured intracellularly or extracellularly, such as IFN-gamma, using flow cytometry. Specific CD4 or CD8 T cell responses to epitopes encoded in the vaccines can be monitored from PBMCs by measuring T cell populations expressing T cell receptors specific for epitope/MHC class I complexes using MHC multimer staining. Specific CD4 or CD8 T cell responses to epitopes encoded in the vaccines can be monitored from PBMCs by measuring the ex vivo expansion of T cell populations following ³H-thymidine, bromodeoxyuridine and carboxyfluoresceine-diacetate– succinimidylester (CFSE) incorporation. The antigen recognition capacity and lytic activity of PBMC-derived T cells that are specific for epitopes encoded in vaccines can be assessed functionally by chromium release assay or alternative colorimetric cytotoxicity assays.

[00399] B cell responses can be measured using one or more methods known in the art such as assays used to determine B cell differentiation (*e.g.*, differentiation into plasma cells), B cell or plasma cell proliferation, B cell or plasma cell activation (*e.g.*, upregulation of costimulatory markers such as CD80 or CD86), antibody class switching, and/or antibody production (*e.g.*, an ELISA).

VII. Antigen Identification

VII.A. Antigen Candidate Identification

[00400] Research methods for NGS analysis transcriptomes have been described and applied in the antigen identification space.^{6,14,15} Certain optimizations for greater sensitivity and specificity for antigen identification in the clinical setting can be considered. These optimizations can be grouped into two areas, those related to laboratory processes and those related to the NGS data analysis. Examples of optimizations are known to those skilled in the art, for example the methods described in more detail in international patent application

publications WO/2017/106638, WO/2018/195357, and WO/2018/208856, each herein incorporated by reference, in their entirety, for all purposes.

[00401]

VII.B. Isolation and Detection of HLA Peptides

[00402] Isolation of HLA-peptide molecules was performed using classic immunoprecipitation (IP) methods after lysis and solubilization of the tissue sample (55-58). A clarified lysate was used for HLA specific IP.

[00403] Immunoprecipitation was performed using antibodies coupled to beads where the antibody is specific for HLA molecules. For a pan-Class I HLA immunoprecipitation, a pan-Class I CR antibody is used, for Class II HLA – DR, an HLA-DR antibody is used. Antibody is covalently attached to NHS-sepharose beads during overnight incubation. After covalent attachment, the beads were washed and aliquoted for IP. (59, 60) Immunoprecipitations can also be performed with antibodies that are not covalently attached to beads. Typically this is done using sepharose or magnetic beads coated with Protein A and/or Protein G to hold the antibody to the column. Some antibodies that can be used to selectively enrich MHC/peptide complex are listed below.

Antibody Name	Specificity
W6/32	Class I HLA-A, B, C
L243	Class II – HLA-DR
Tu36	Class II – HLA-DR
LN3	Class II – HLA-DR
Tu39	Class II – HLA-DR, DP, DQ
SPVL3	Class II – HLA - DQ
B7/21	Class II – HLA - DP

[00404] The clarified tissue lysate is added to the antibody beads for the immunoprecipitation. After immunoprecipitation, the beads are removed from the lysate and the lysate stored for additional experiments, including additional IPs. The IP beads are washed to remove non-specific binding and the HLA/peptide complex is eluted from the beads using standard techniques. The protein components are removed from the peptides using a molecular weight spin column or C18 fractionation. The resultant peptides are taken to dryness by SpeedVac evaporation and in some instances are stored at -20C prior to MS analysis. HLA IPs

can also be performed in 96well plate format using plates that contain filter bottoms. Use of the plates allows for multiple IPs to be performed in tandem.

[00405] Dried peptides are reconstituted in an HPLC buffer suitable for reverse phase chromatography and loaded onto a C-18 microcapillary HPLC column for gradient elution in a Fusion Lumos mass spectrometer (Thermo). MS1 spectra of peptide mass/charge (m/z) were collected in the Orbitrap detector at high resolution followed by MS2 low resolution scans collected in the ion trap detector after HCD fragmentation of the selected ion. Additionally, MS2 spectra can be obtained using either CID or ETD fragmentation methods or any combination of the three techniques to attain greater amino acid coverage of the peptide. MS2 spectra can also be measured with high resolution mass accuracy in the Orbitrap detector.

[00406] MS2 spectra from each analysis are searched against a protein database using Comet (61, 62) and the peptide identification are scored using Percolator (63-65). Additional sequencing is performed using PEAKS studio (Bioinformatics Solutions Inc.) and other search engines or sequencing methods can be used including spectral matching and *de novo* sequencing (97).

VIII. Presentation Model

[00407] Presentation models can be used to identify likelihoods of peptide presentation in patients. Various presentation models are known to those skilled in the art, for example the presentation models described in more detail in US Pat No. 10,055,540, US Application Pub. No. US20200010849A1 and US20110293637, and international patent application publications WO/2018/195357, WO/2018/208856, WO2016187508, each herein incorporated by reference, in their entirety, for all purposes.

IX. Training Module

[00408] Training modules can be used to construct one or more presentation models based on training data sets that generate likelihoods of whether peptide sequences will be presented by MHC alleles associated with the peptide sequences. Various training modules are known to those skilled in the art, for example the presentation models described in more detail in US Pat No. 10,055,540, US Application Pub. No. US20200010849A1, and international patent application publication WO/2018/195357, and WO/2018/208856, each herein incorporated by reference, in their entirety, for all purposes. A training module can construct a presentation model to predict presentation likelihoods of peptides on a per-allele basis. A training module

can also construct a presentation model to predict presentation likelihoods of peptides in a multiple-allele setting where two or more MHC alleles are present.

X. Prediction Module

[00409] A prediction module can be used to receive sequence data and select candidate antigens in the sequence data using a presentation model. Specifically, the sequence data may be DNA sequences, RNA sequences, and/or protein sequences extracted from infected cells of patients. A prediction module may identify candidate antigens that are pathogen-derived peptides, a virally-derived peptides, a bacterially-derived peptides, a fungally-derived peptides, and a parasitically-derived peptides by comparing sequence data extracted from normal tissue cells of a patient with the sequence data extracted from infected cells of the patient to identify portions containing one or more infectious disease organism associated antigens. A prediction module may identify candidate antigens that have altered expression in an infected cell or infected tissue in comparison to a normal cell or tissue by comparing sequence data extracted from normal tissue cells of a patient with the sequence data extracted from infected tissue cells of the patient to identify improperly expressed candidate antigens (*e.g.*, identifying expressed polynucleotides and/or polypeptides specific to an infectious disease).

[00410] A presentation module can apply one or more presentation model to processed peptide sequences to estimate presentation likelihoods of the peptide sequences. Specifically, the prediction module may select one or more candidate antigen peptide sequences that are likely to be presented on infected cell HLA molecules by applying presentation models to the candidate antigens. In one implementation, the presentation module selects candidate antigen sequences that have estimated presentation likelihoods above a predetermined threshold. In another implementation, the presentation model selects the N candidate antigen sequences that have the highest estimated presentation likelihoods (where N is generally the maximum number of epitopes that can be delivered in a vaccine). A vaccine including the selected candidate antigens for a given patient can be injected into the patient to induce immune responses.

XI.B.Cassette Design Module

XI.B.1 Overview

[00411] A cassette design module can be used to generate a vaccine cassette sequence based on selected candidate peptides for injection into a patient. For example, a cassette design module can be used to generate a sequence encoding concatenated epitope sequences, such as concatenated T cell epitopes. Various cassette design modules are known to those

skilled in the art, for example the cassette design modules described in more detail in US Pat No. 10,055,540, US Application Pub. No. US20200010849A1, and international patent application publications WO/2018/195357 and WO/2018/208856, each herein incorporated by reference, in their entirety, for all purposes.

[00412] A set of therapeutic epitopes may be generated based on the selected peptides determined by a prediction module associated with presentation likelihoods above a predetermined threshold, where the presentation likelihoods are determined by the presentation models. However it is appreciated that in other embodiments, the set of therapeutic epitopes may be generated based on any one or more of a number of methods (alone or in combination), for example, based on binding affinity or predicted binding affinity to HLA class I or class II alleles of the patient, binding stability or predicted binding stability to HLA class I or class II alleles of the patient, random sampling, and the like.

[00413] Therapeutic epitopes may correspond to selected peptides themselves. Therapeutic epitopes may also include C- and/or N-terminal flanking sequences in addition to the selected peptides. N- and C-terminal flanking sequences can be the native N- and C-terminal flanking sequences of the therapeutic vaccine epitope in the context of its source protein. Therapeutic epitopes can represent a fixed-length epitope. Therapeutic epitopes can represent a variable-length epitope, in which the length of the epitope can be varied depending on, for example, the length of the C- or N-flanking sequence. For example, the C-terminal flanking sequence and the N-terminal flanking sequence can each have varying lengths of 2-5 residues, resulting in 16 possible choices for the epitope.

[00414] A cassette design module can also generate cassette sequences by taking into account presentation of junction epitopes that span the junction between a pair of therapeutic epitopes in the cassette. Junction epitopes are novel non-self but irrelevant epitope sequences that arise in the cassette due to the process of concatenating therapeutic epitopes and linker sequences in the cassette. The novel sequences of junction epitopes are different from the therapeutic epitopes of the cassette themselves.

[00415] A cassette design module can generate a cassette sequence that reduces the likelihood that junction epitopes are presented in the patient. Specifically, when the cassette is injected into the patient, junction epitopes have the potential to be presented by HLA class I or HLA class II alleles of the patient, and stimulate a CD8 or CD4 T-cell response, respectively. Such reactions are often times undesirable because T-cells reactive to the

junction epitopes have no therapeutic benefit, and may diminish the immune response to the selected therapeutic epitopes in the cassette by antigenic competition.⁷⁶

[00416] A cassette design module can iterate through one or more candidate cassettes, and determine a cassette sequence for which a presentation score of junction epitopes associated with that cassette sequence is below a numerical threshold. The junction epitope presentation score is a quantity associated with presentation likelihoods of the junction epitopes in the cassette, and a higher value of the junction epitope presentation score indicates a higher likelihood that junction epitopes of the cassette will be presented by HLA class I or HLA class II or both.

[00417] In one embodiment, a cassette design module may determine a cassette sequence associated with the lowest junction epitope presentation score among the candidate cassette sequences.

[00418] A cassette design module may iterate through one or more candidate cassette sequences, determine the junction epitope presentation score for the candidate cassettes, and identify an optimal cassette sequence associated with a junction epitope presentation score below the threshold.

[00419] A cassette design module may further check the one or more candidate cassette sequences to identify if any of the junction epitopes in the candidate cassette sequences are self-epitopes for a given patient for whom the vaccine is being designed. To accomplish this, the cassette design module checks the junction epitopes against a known database such as BLAST. In one embodiment, the cassette design module may be configured to design cassettes that avoid junction self-epitopes.

[00420] A cassette design module can perform a brute force approach and iterate through all or most possible candidate cassette sequences to select the sequence with the smallest junction epitope presentation score. However, the number of such candidate cassettes can be prohibitively large as the capacity of the vaccine increases. For example, for a vaccine capacity of 20 epitopes, the cassette design module has to iterate through $\sim 10^{18}$ possible candidate cassettes to determine the cassette with the lowest junction epitope presentation score. This determination may be computationally burdensome (in terms of computational processing resources required), and sometimes intractable, for the cassette design module to complete within a reasonable amount of time to generate the vaccine for the patient. Moreover, accounting for the possible junction epitopes for each candidate cassette can be even more burdensome. Thus, a cassette design module may select a cassette sequence

based on ways of iterating through a number of candidate cassette sequences that are significantly smaller than the number of candidate cassette sequences for the brute force approach.

[00421] A cassette design module can generate a subset of randomly or at least pseudo-randomly generated candidate cassettes, and selects the candidate cassette associated with a junction epitope presentation score below a predetermined threshold as the cassette sequence. Additionally, the cassette design module may select the candidate cassette from the subset with the lowest junction epitope presentation score as the cassette sequence. For example, the cassette design module may generate a subset of ~1 million candidate cassettes for a set of 20 selected epitopes, and select the candidate cassette with the smallest junction epitope presentation score. Although generating a subset of random cassette sequences and selecting a cassette sequence with a low junction epitope presentation score out of the subset may be sub-optimal relative to the brute force approach, it requires significantly less computational resources thereby making its implementation technically feasible. Further, performing the brute force method as opposed to this more efficient technique may only result in a minor or even negligible improvement in junction epitope presentation score, thus making it not worthwhile from a resource allocation perspective. A cassette design module can determine an improved cassette configuration by formulating the epitope sequence for the cassette as an asymmetric traveling salesman problem (TSP). Given a list of nodes and distances between each pair of nodes, the TSP determines a sequence of nodes associated with the shortest total distance to visit each node exactly once and return to the original node. For example, given cities A, B, and C with known distances between each other, the solution of the TSP generates a closed sequence of cities, for which the total distance traveled to visit each city exactly once is the smallest among possible routes. The asymmetric version of the TSP determines the optimal sequence of nodes when the distance between a pair of nodes are asymmetric. For example, the “distance” for traveling from node A to node B may be different from the “distance” for traveling from node B to node A. By solving for an improved optimal cassette using an asymmetric TSP, the cassette design module can find a cassette sequence that results in a reduced presentation score across the junctions between epitopes of the cassette. The solution of the asymmetric TSP indicates a sequence of therapeutic epitopes that correspond to the order in which the epitopes should be concatenated in a cassette to minimize the junction epitope presentation score across the junctions of the cassette. A cassette sequence determined through this approach can result in

a sequence with significantly less presentation of junction epitopes while potentially requiring significantly less computational resources than the random sampling approach, especially when the number of generated candidate cassette sequences is large. Illustrative examples of different computational approaches and comparisons for optimizing cassette design are described in more detail in US Pat No. 10,055,540, US Application Pub. No. US20200010849A1, and international patent application publications WO/2018/195357, and WO/2018/208856, each herein incorporated by reference, in their entirety, for all purposes.

[00422] A cassette design module can also generate cassette sequences by taking into account additional protein sequences encoded in the vaccine. For example, a cassette design module used to generate a sequence encoding concatenated T cell epitopes can take into account T cell epitopes already encoded by additional protein sequences present in the vaccine (*e.g.*, full-length protein sequences), such as by removing T cell epitopes already encoded by the additional protein sequences from the list of candidate sequences.

[00423] A cassette design module can also generate cassette sequences by taking into account the size of the sequences. Without wishing to be bound by theory, in general, increased cassette size can negatively impact vaccine aspects, such as vaccine production and/or vaccine efficacy. In one example, the cassette design module can take into account overlapping sequences, such as overlapping T cell epitope sequences. In general, a single sequence containing overlapping T cell epitope sequences (also referred to as a “frame”) is more efficient than separately linking individual T cell epitope sequences as it reduces the sequence size needed to encode the multiple peptides. Accordingly, in an illustrative example, a cassette design module used to generate a sequence encoding concatenated T cell epitopes can take into account the cost/benefit of extending a candidate T cell epitope to encode one or more additional T cell epitopes, such as determining the benefit gained in additional population coverage for an MHC presenting the additional T cell epitope versus the cost of increasing the size of the sequence.

[00424] A cassette design module can also generate cassette sequences by taking into account other aspects that improve potential safety, such as limiting encoding or the potential to encode a functional protein, functional protein domain, functional protein subunit, or functional protein fragment potentially presentially presenting a safety risk. In some cases, a cassette design module can limit sequence size of encoded peptides such that they are less than 50%, less than 49%, less than 48%, less than 47%, less than 46%, less than 45%, less than 45%, less than 43%, less than 42%, less than 41%, less than 40%, less than 39%, less than

38%, less than 37%, less than 36%, less than 35%, less than 34%, or less than 33% of the translated, corresponding full-length protein. In some cases, a cassette design module can limit sequence size of encoded peptides such that a single contiguous sequence is less than 50% of the translated, corresponding full-length protein, but more than one sequence may be derived from the same translated, corresponding full-length protein and together encode more than 50%. In an illustrative example, if a single sequence containing overlapping T cell epitope sequences (“frame”) is larger than 50% of the translated, corresponding full-length protein, the frame can be split into multiple frames (*e.g.*, f1, f2 etc.) such that each frame is less than 50% of the translated, corresponding full-length protein. A cassette design module can also limit sequence size of encoded peptides such that a single contiguous sequence is less than 49%, less than 48%, less than 47%, less than 46%, less than 45%, less than 45%, less than 43%, less than 42%, less than 41%, less than 40%, less than 39%, less than 38%, less than 37%, less than 36%, less than 35%, less than 34%, or less than 33% of the translated, corresponding full-length protein. Where multiple frames from the same gene are encoded, the multiple frames can have overlapping sequences with each other, in other words each separately encode the same sequence. Where multiple frames from the same gene are encoded, the two or more nucleic acid sequences derived from the same gene can be ordered such that a first nucleic acid sequence cannot be immediately followed by or linked to a second nucleic acid sequence if the second nucleic acid sequence follows, immediately or not, the first nucleic acid sequence in the corresponding gene. . For example, if there are 3 frames within the same gene (f1,f2,f3 in increasing order of amino acid position):

- The following cassette orderings are not allowed:
 - f1 immediately followed by f2
 - f2 immediately followed by f3
 - f1 immediately followed by f3
- The following cassette orderings are allowed:
 - f3 immediately followed by f2

[00425] f2 immediately followed by f1

XII. Example Computer

[00426] A computer can be used for any of the computational methods described herein. One skilled in the art will recognize a computer can have different architectures. Examples of computers are known to those skilled in the art, for example the computers described in more

detail in US Pat No. 10,055,540, US Application Pub. No. US20200010849A1, and international patent application publications WO/2018/195357, and WO/2018/208856, each herein incorporated by reference, in their entirety, for all purposes.

[00427]

XIII. Antigen Delivery Vector Example

[00428] Below are examples of specific embodiments for carrying out the present invention. The examples are offered for illustrative purposes only, and are not intended to limit the scope of the present invention in any way. Efforts have been made to ensure accuracy with respect to numbers used (e.g., amounts, temperatures, etc.), but some experimental error and deviation should, of course, be allowed for.

[00429] The practice of the present invention will employ, unless otherwise indicated, conventional methods of protein chemistry, biochemistry, recombinant DNA techniques and pharmacology, within the skill of the art. Such techniques are explained fully in the literature. See, e.g., T.E. Creighton, *Proteins: Structures and Molecular Properties* (W.H. Freeman and Company, 1993); A.L. Lehninger, *Biochemistry* (Worth Publishers, Inc., current addition); Sambrook, et al., *Molecular Cloning: A Laboratory Manual* (2nd Edition, 1989); *Methods In Enzymology* (S. Colowick and N. Kaplan eds., Academic Press, Inc.); *Remington's Pharmaceutical Sciences*, 18th Edition (Easton, Pennsylvania: Mack Publishing Company, 1990); Carey and Sundberg *Advanced Organic Chemistry 3rd Ed.* (Plenum Press) Vols A and B(1992).

XIII.A. Antigen Cassette Design

[00430] Through vaccination, multiple class I MHC restricted antigens that stimulate the corresponding cellular immune response(s) can be delivered. In one example, a vaccine cassette was engineered to encode multiple epitopes as a single gene product where the epitopes were either embedded within their natural, surrounding peptide sequence or spaced by non-natural linker sequences. Several design parameters were identified that could potentially impact antigen processing and presentation and therefore the magnitude and breadth of the antigen specific CD8 T cell responses. In the present example, several model cassettes were designed and constructed to evaluate: (1) whether robust T cell responses could be generated to multiple epitopes incorporated in a single expression cassette; (2) what makes an optimal linker placed between the antigens within the expression cassette- that leads to optimal processing and presentation of all epitopes; (3) if the relative position of the epitopes within the cassette

impact T cell responses; (4) whether the number of epitopes within a cassette influences the magnitude or quality of the T cell responses to individual epitopes; (5) if the addition of cellular targeting sequences improves T cell responses.

[00431] Two readouts were developed to evaluate antigen presentation and T cell responses specific for marker epitopes within the model cassettes: (1) an *in vitro* cell-based screen which allowed assessment of antigen presentation as gauged by the activation of specially engineered reporter T cells (Aarnoudse et al., 2002; Nagai et al., 2012); and (2) an *in vivo* assay that used HLA-A2 transgenic mice (Vitiello et al., 1991) to assess post-vaccination immunogenicity of cassette-derived epitopes of human origin by their corresponding epitope-specific T cell responses (Cornet et al., 2006; Depla et al., 2008; Ishioka et al., 1999).

XIV. ChAd Antigen Cassette Delivery Vector

XIV.A. ChAd Antigen Cassette Delivery Vector Construction

[00432] In one example, Chimpanzee adenovirus (ChAd) was engineered to be a delivery vector for antigen cassettes. In a further example, a full-length ChAdV68 vector was synthesized based on AC_000011.1 (sequence 2 from Patent US 6083716) with E1 (nt 457 to 3014) and E3 (nt 27,816- 31,332) sequences deleted. Reporter genes under the control of the CMV promoter/enhancer were inserted in place of the deleted E1 sequences. Transfection of this clone into HEK293 cells did not yield infectious virus. To confirm the sequence of the wild-type C68 virus, isolate VR-594 was obtained from the ATCC, passaged, and then independently sequenced (SEQ ID NO:10). When comparing the AC_000011.1 sequence to the ATCC VR-594 sequence (SEQ ID NO:10) of wild-type ChAdV68 virus, 6 nucleotide differences were identified. In one example, a modified ChAdV68 vector was generated based on AC_000011.1, with the corresponding ATCC VR-594 nucleotides substituted at five positions (ChAdV68.5WTnt SEQ ID NO:1).

[00433] In another example, a modified ChAdV68 vector was generated based on AC_000011.1 with E1 (nt 577 to 3403) and E3 (nt 27,816- 31,332) sequences deleted and the corresponding ATCC VR-594 nucleotides substituted at four positions. A GFP reporter (ChAdV68.4WTnt.GFP; SEQ ID NO:11) or model neoantigen cassette (ChAdV68.4WTnt.MAG25mer; SEQ ID NO:12) under the control of the CMV promoter/enhancer was inserted in place of deleted E1 sequences.

[00434] In another example, a modified ChAdV68 vector was generated based on AC_000011.1 with E1 (nt 577 to 3403) and E3 (nt 27,125- 31,825) sequences deleted and the corresponding ATCC VR-594 nucleotides substituted at five positions. A GFP reporter (ChAdV68.5WTnt.GFP; SEQ ID NO:13) or model neoantigen cassette (ChAdV68.5WTnt.MAG25mer; SEQ ID NO:2) under the control of the CMV promoter/enhancer was inserted in place of deleted E1 sequences.

[00435] Relevant vectors are described below:

- Full-Length ChAdVC68 sequence “ChAdV68.5WTnt” (SEQ ID NO:1); AC_000011.1 sequence with corresponding ATCC VR-594 nucleotides substituted at five positions.
- ATCC VR-594 C68 (SEQ ID NO:10); Independently sequenced; Full-Length C68

- ChAdV68.4WTnt.GFP (SEQ ID NO:11); AC_000011.1 with E1 (nt 577 to 3403) and E3 (nt 27,816- 31,332) sequences deleted; corresponding ATCC VR-594 nucleotides substituted at four positions; GFP reporter under the control of the CMV promoter/enhancer inserted in place of deleted E1
- ChAdV68.4WTnt.MAG25mer (SEQ ID NO:12); AC_000011.1 with E1 (nt 577 to 3403) and E3 (nt 27,816- 31,332) sequences deleted; corresponding ATCC VR-594 nucleotides substituted at four positions; model neoantigen cassette under the control of the CMV promoter/enhancer inserted in place of deleted E1
- ChAdV68.5WTnt.GFP (SEQ ID NO:13); AC_000011.1 with E1 (nt 577 to 3403) and E3 (nt 27,125- 31,825) sequences deleted; corresponding ATCC VR-594 nucleotides substituted at five positions; GFP reporter under the control of the CMV promoter/enhancer inserted in place of deleted E1

XV. Alphavirus Antigen Cassette Delivery Vector

XV.B. Alphavirus Vector

[00436] In one implementation of the present invention, a RNA alphavirus backbone for the antigen expression system was generated from a Venezuelan Equine Encephalitis (VEE) (Kinney, 1986, Virology 152: 400-413) based self-replicating RNA (srRNA) vector. In one example, the sequences encoding the structural proteins of VEE located 3' of the 26S sub-genomic promoter were deleted (VEE sequences 7544 to 11,175 deleted; numbering based on Kinney *et al* 1986; SEQ ID NO:6) and replaced by antigen sequences (SEQ ID NO:14 and SEQ ID NO:4) or a luciferase reporter (e.g., VEE-Luciferase, SEQ ID NO:15).

XVI. Non-Human Primate Studies

[00437] Various dosing protocols using ChAdV68 and self-replicating RNA (srRNA) were evaluated in non-human primates (NHP).

Materials and Methods

[00438] A priming vaccine was injected intramuscularly (IM) in each NHP to initiate the study (vaccine prime). One or more boosting vaccines (vaccine boost) were also injected intramuscularly in each NHP. Bilateral injections per dose were administered according to groups outlined in tables and summarized below.

Immunizations

[00439] Mamu-A*01 Indian rhesus macaques were immunized bilaterally with 1×10^{12} viral particles (5×10^{11} viral particles per injection) of ChAdV68.5WTnt.MAG25mer, 30 ug of VEE-MAG25MER srRNA, 100 ug of VEE-MAG25mer srRNA or 300 ug of VEE-MAG25mer srRNA formulated in LNP-1 or LNP-2. Vaccine boosts of 30 ug, 100 ug or 300 ug VEE-MAG25mer srRNA were administered intramuscularly at the indicated time after prime vaccination.

Immune Monitoring

[00440] PBMCs were isolated at indicated times after prime vaccination using Lymphocyte Separation Medium (LSM, MP Biomedicals) and LeucoSep separation tubes (Greiner Bio-One) and resuspended in RPMI containing 10% FBS and penicillin/streptomycin. Cells were counted on the Attune NxT flow cytometer (Thermo Fisher) using propidium iodide staining to exclude dead and apoptotic cells. Cells were then adjusted to the appropriate concentration of live cells for subsequent analysis. For each monkey in the studies, T cell responses were measured using ELISpot or flow cytometry methods. T cell responses to 6 different rhesus macaque Mamu-A*01 class I epitopes encoded in the vaccines were monitored from PBMCs by measuring induction of cytokines, such as IFN-gamma, using ex vivo enzyme-linked immunospot (ELISpot) analysis. ELISpot analysis was performed according to ELISpot harmonization guidelines {DOI: 10.1038/nprot.2015.068} with the monkey IFNg ELISpotPLUS kit (MABTECH). 200,000 PBMCs were incubated with 10uM of the indicated peptides for 16 hours in 96-well IFNg antibody coated plates. Spots were developed using alkaline phosphatase. The reaction was timed for 10 minutes and was terminated by running plate under tap water. Spots were counted using an AID vSpot Reader Spectrum. For ELISpot analysis, wells with saturation >50% were recorded as “too numerous to count”. Samples with deviation of replicate wells > 10% were excluded from analysis. Spot counts were then corrected for well confluency using the formula: $\text{spot count} + 2 \times (\text{spot count} \times \% \text{confluence} / [100\% - \% \text{confluence}])$. Negative background was corrected by subtraction of spot counts in the negative peptide stimulation wells from the antigen stimulated wells. Finally, wells labeled too numerous to count were set to the highest observed corrected value, rounded up to the nearest hundred.

[00441] Specific CD4 and CD8 T cell responses to 6 different rhesus macaque Mamu-A*01 class I epitopes encoded in the vaccines were monitored from PBMCs by measuring induction of intracellular cytokines, such as IFN-gamma, using flow cytometry. The results from both methods indicate that cytokines were induced in an antigen-specific manner to epitopes.

Immunogenicity in rhesus macaques

[00442] This study was designed to (a) evaluate the immunogenicity and preliminary safety of VEE-MAG25mer srRNA 30 µg and 100 µg doses as a homologous prime/boost or heterologous prime/boost in combination with ChAdV68.5WTnt.MAG25mer; (b) compare the immune responses of VEE-MAG25mer srRNA in lipid nanoparticles using LNP1 versus LNP2; (c) evaluate the kinetics of T-cell responses to VEE-MAG25mer srRNA and ChAdV68.5WTnt.MAG25mer immunizations.

[00443] The study arm was conducted in Mamu-A*01 Indian rhesus macaques to demonstrate immunogenicity. Select antigens used in this study are only recognized in Rhesus macaques, specifically those with a Mamu-A*01 MHC class I haplotype. Mamu-A*01 Indian rhesus macaques were randomized to the different study arms (6 macaques per group) and administered an IM injection bilaterally with either ChAdV68.5WTnt.MAG25mer or VEE-MAG25mer srRNA vector encoding model antigens that includes multiple Mamu-A*01 restricted epitopes. The study arms were as described below.

[00444] **Table A:** Non-GLP immunogenicity study in Indian Rhesus Macaques

Group	Prime	Boost 1	Boost 2
1	VEE-MAG25mer srRNA -LNP1(30 µg)	VEE-MAG25mer srRNA -LNP1 (30 µg)	VEE-MAG25mer srRNA -LNP1 (30 µg)
2	VEE-MAG25mer srRNA -LNP1 (100 µg)	VEE-MAG25mer srRNA -LNP1 (100 µg)	VEE-MAG25mer srRNA -LNP1 (100 µg)
3	VEE-MAG25mer srRNA -LNP2 (100 µg)	VEE-MAG25mer srRNA -LNP2 (100 µg)	VEE-MAG25mer srRNA -LNP2 (100 µg)
4	ChAdV68.5WTnt. MAG25mer	VEE-MAG25mer srRNA -LNP1 (100 µg)	VEE-MAG25mer srRNA -LNP1 (100 µg)

[00445] PBMCs were collected prior to immunization and on weeks 1, 2, 3, 4, 5, 6, 8, 9, and 10 after the initial immunization for immune monitoring.

Results

[00446] Antigen-specific cellular immune responses in peripheral blood mononuclear cells (PBMCs) were measured to six different Mamu-A*01 restricted epitopes prior to immunization and 1, 2, 3, 4, 5, 6, 8, 9, and 10 weeks after the initial immunization. Animals received a boost immunization with VEE-MAG25mer srRNA on weeks 4 and 8 with either 30 µg or 100 µg doses, and either formulated with LNP1 or LNP2, as described in Table A. Combined immune responses to all six epitopes were plotted for each immune monitoring timepoint (Fig. 1A-D and Tables B-E).

[00447] Combined antigen-specific immune responses were observed at all measurements with 170, 14, 15, 11, 7, 8, 14, 17, 12 SFCs per 10^6 PBMCs (six epitopes combined) 1, 2, 3, 4, 5, 6, 8, 9, or 10 weeks after an initial VEE-MAG25mer srRNA-LNP1(30 µg) prime immunization, respectively (FIG. 1A). Combined antigen-specific immune responses were observed at all measurements with 108, -3, 14, 1, 37, 4, 105, 17, 25 SFCs per 10^6 PBMCs (six epitopes combined) 1, 2, 3, 4, 5, 6, 8, 9, or 10 weeks after an initial VEE-MAG25mer srRNA-LNP1(100 µg) prime immunization, respectively (FIG. 1B). Combined antigen-specific immune responses were observed at all measurements with -17, 38, 14, -2, 87, 21, 104, 129, 89 SFCs per 10^6 PBMCs (six epitopes combined) 1, 2, 3, 4, 5, 6, 8, 9, or 10 weeks after an initial VEE-MAG25mer srRNA-LNP2(100 µg) prime immunization, respectively (FIG. 1C). Negative values are a result of normalization to pre-bleed values for each epitope/animal.

[00448] Combined antigen-specific immune responses were observed at all measurements with 1218, 1784, 1866, 973, 1813, 747, 797, 1249, and 547 SFCs per 10^6 PBMCs (six epitopes combined) 1, 2, 3, 4, 5, 6, 8, 9, or 10 weeks after an initial ChAdV68.5WTnt.MAG25mer prime immunization, respectively (FIG. 1D). The immune response showed the expected profile with peak immune responses measured ~2-3 weeks after the prime immunization followed by a contraction in the immune response after 4 weeks. Combined antigen-specific cellular immune responses of 1813 SFCs per 10^6 PBMCs (six epitopes combined) were measured 5 weeks after the initial immunization with ChAdV68.5WTnt.MAG25mer (i.e., 1 week after the first boost with VEE-MAG25mer srRNA). The immune response measured 1 week after the first boost with VEE-MAG25mer srRNA (week 5) was comparable to the peak immune response measured for the ChAdV68.5WTnt.MAG25mer prime immunization (week 3) (FIG. 20D). Combined antigen-specific cellular immune responses of 1249 SFCs per 10^6 PBMCs (six epitopes combined) was measured 9 weeks after the initial immunization with ChAdV68.5WTnt.MAG25mer, respectively (i.e., 1 week after the second boost with VEE-

MAG25mer srRNA). The immune responses measured 1 week after the second boost with VEE-MAG25mer srRNA (week 9) was ~2-fold higher than that measured just before the boost immunization (FIG. 1D).

[00449] Table B: Mean spot forming cells (SFC) per 10^6 PBMCs for each epitope $\pm$ SEM for VEE-MAG25mer srRNA-LNP1(30 μ g) (Group 1)

Wk	Antigen					
	Env CL9	Env TL9	Gag CM9	Gag LW9	Pol SV9	Tat TL8
1	0 $\pm$ 0	0 $\pm$ 0	0 $\pm$ 0	0 $\pm$ 0	0 $\pm$ 0	0 $\pm$ 0
2	39.7 $\pm$ 22.7	35.4 $\pm$ 25.1	3.2 $\pm$ 3.6	33 $\pm$ 28.1	30.9 $\pm$ 20.3	28.3 $\pm$ 17.5
3	2 $\pm$ 2.4	0.2 $\pm$ 1.8	1.8 $\pm$ 2.4	3.7 $\pm$ 1.9	1.7 $\pm$ 2.8	4.9 $\pm$ 2.3
4	1 $\pm$ 1.8	0.3 $\pm$ 1.2	5.5 $\pm$ 3.6	2.3 $\pm$ 2.2	5.7 $\pm$ 2.7	0.8 $\pm$ 0.8
5	0.5 $\pm$ 0.9	1.4 $\pm$ 3.8	3.1 $\pm$ 1.6	2.3 $\pm$ 2.7	1.9 $\pm$ 2	1.4 $\pm$ 1.2
6	1.9 $\pm$ 1.8	-0.3 $\pm$ 3	1.7 $\pm$ 1.2	1.4 $\pm$ 1.4	0.8 $\pm$ 1.1	1.1 $\pm$ 1
8	-0.4 $\pm$ 0.8	-0.9 $\pm$ 2.9	0.5 $\pm$ 1.3	3 $\pm$ 1.1	2.2 $\pm$ 2.1	3.7 $\pm$ 2
9	1 $\pm$ 1.7	1.2 $\pm$ 4.2	7.2 $\pm$ 3.9	0.5 $\pm$ 0.7	1.6 $\pm$ 3	3 $\pm$ 1
10	3.8 $\pm$ 1.8	11 $\pm$ 5	-1.1 $\pm$ 1.1	1.9 $\pm$ 0.9	1.3 $\pm$ 1.6	0.2 $\pm$ 0.5

[00450] Table C: Mean spot forming cells (SFC) per 10^6 PBMCs for each epitope $\pm$ SEM for VEE-MAG25mer srRNA-LNP1(100 μ g) (Group 2)

Wk	Antigen					
	Env CL9	Env TL9	Gag CM9	Gag LW9	Pol SV9	Tat TL8
1	0 $\pm$ 0	0 $\pm$ 0	0 $\pm$ 0	0 $\pm$ 0	0 $\pm$ 0	0 $\pm$ 0
2	7.9 $\pm$ 17.2	23.2 $\pm$ 17.4	11.4 $\pm$ 4.9	41.7 $\pm$ 16.5	15 $\pm$ 13.5	8.9 $\pm$ 6.2
3	-3.1 $\pm$ 4.6	-7.2 $\pm$ 6.5	2.3 $\pm$ 2.3	-0.3 $\pm$ 2.7	2.7 $\pm$ 5.1	2.2 $\pm$ 1.4
4	1.9 $\pm$ 3.8	-6.2 $\pm$ 7.6	10.5 $\pm$ 4.1	1.2 $\pm$ 2.9	5.6 $\pm$ 4.9	1.1 $\pm$ 0.8
5	-2.6 $\pm$ 7	-8 $\pm$ 5.9	1.5 $\pm$ 1.7	6.4 $\pm$ 2.3	0.7 $\pm$ 4.3	3.3 $\pm$ 1.3
6	6.3 $\pm$ 6.3	4.4 $\pm$ 8.3	6.6 $\pm$ 4.4	5.2 $\pm$ 5.2	3.9 $\pm$ 5	10.8 $\pm$ 6.9
8	-3.6 $\pm$ 7.2	-6.8 $\pm$ 7.3	-0.8 $\pm$ 1.2	3.4 $\pm$ 4.2	6.4 $\pm$ 7.5	5.7 $\pm$ 2.7
9	8.1 $\pm$ 2.4	20.6 $\pm$ 23.4	18.9 $\pm$ 5.7	8.1 $\pm$ 8.9	9 $\pm$ 11.2	40 $\pm$ 17.6
10	3.1 $\pm$ 8	-3.9 $\pm$ 8.5	3.3 $\pm$ 1.8	0.6 $\pm$ 2.9	7.4 $\pm$ 6.4	6.1 $\pm$ 2.5

[00451] Table D: Mean spot forming cells (SFC) per 10^6 PBMCs for each epitope $\pm$ SEM for VEE-MAG25mer srRNA-LNP2(100 μ g) (Group 3)

Wk	Antigen					
	Env CL9	Env TL9	Gag CM9	Gag LW9	Pol SV9	Tat TL8
1	0 $\pm$ 0	0 $\pm$ 0	0 $\pm$ 0	0 $\pm$ 0	0 $\pm$ 0	0 $\pm$ 0
2	-5.9 $\pm$ 3.8	-0.3 $\pm$ 0.5	-0.5 $\pm$ 1.5	-5.7 $\pm$ 6.1	-1 $\pm$ 1.3	-3.2 $\pm$ 5.5
3	0.7 $\pm$ 5.2	3.4 $\pm$ 2.4	4.2 $\pm$ 4.6	18.3 $\pm$ 15.5	11.9 $\pm$ 5.1	-0.4 $\pm$ 8.2
4	-3.8 $\pm$ 5.5	2.3 $\pm$ 1.8	11.3 $\pm$ 6.1	-3.1 $\pm$ 5.6	8.5 $\pm$ 4	-1.5 $\pm$ 6.1
5	-3.7 $\pm$ 5.7	-0.1 $\pm$ 0.7	-0.2 $\pm$ 1.6	3.4 $\pm$ 8.5	3 $\pm$ 3.1	-4.6 $\pm$ 5
6	12.3 $\pm$ 15	7.8 $\pm$ 4.9	24.7 $\pm$ 19.8	23.2 $\pm$ 22.5	18.7 $\pm$ 15.8	0.5 $\pm$ 6.2
8	5.9 $\pm$ 12.3	-0.1 $\pm$ 0.7	-0.5 $\pm$ 1.3	8.8 $\pm$ 14.4	8.7 $\pm$ 8	-1.3 $\pm$ 4
9	16.1 $\pm$ 13.4	16.5 $\pm$ 4	22.9 $\pm$ 4.2	13 $\pm$ 13.2	16.4 $\pm$ 7.8	19.6 $\pm$ 9.2
10	29.9 $\pm$ 21.8	22 $\pm$ 19.5	0.5 $\pm$ 2.6	22.2 $\pm$ 22.6	35.3 $\pm$ 15.8	19.4 $\pm$ 17.3

[00452] Table E: Mean spot forming cells (SFC) per 10^6 PBMCs for each epitope $\pm$ SEM for ChAdV68.5WTnt.MAG25mer prime

Wk	Antigen					
	Env CL9	Env TL9	Gag CM9	Gag LW9	Pol SV9	Tat TL8
1	178 $\pm$ 68.7	206.5 $\pm$ 94.8	221.2 $\pm$ 120	15.4 $\pm$ 16.7	33.3 $\pm$ 25.9	563.5 $\pm$ 174.4
2	311.2 $\pm$ 165.5	278.8 $\pm$ 100.9	344.6 $\pm$ 110.8	46.3 $\pm$ 13.5	181.6 $\pm$ 76.8	621.4 $\pm$ 220.9
3	277.3 $\pm$ 101.1	359.6 $\pm$ 90.5	468.2 $\pm$ 106.6	41.7 $\pm$ 11.1	169.8 $\pm$ 57.8	549.4 $\pm$ 115.7
4	140 $\pm$ 46.5	169.6 $\pm$ 46.8	239.4 $\pm$ 37	26.5 $\pm$ 11.4	75 $\pm$ 31.6	322.2 $\pm$ 50.7
5	155.6 $\pm$ 62.1	406.7 $\pm$ 96.4	542.7 $\pm$ 143.3	35.1 $\pm$ 16.6	134.2 $\pm$ 53.7	538.5 $\pm$ 91.9
6	78.9 $\pm$ 42.5	95.5 $\pm$ 29.4	220.9 $\pm$ 75.3	-1.4 $\pm$ 5.3	43.4 $\pm$ 19.6	308.1 $\pm$ 42.6
8	88.4 $\pm$ 30.4	162.1 $\pm$ 30.3	253.4 $\pm$ 78.6	21.4 $\pm$ 11.2	53.7 $\pm$ 22.3	217.8 $\pm$ 45.2
9	158.5 $\pm$ 69	322.3 $\pm$ 87.2	338.2 $\pm$ 137.1	5.6 $\pm$ 12.4	109.2 $\pm$ 17.9	314.8 $\pm$ 43.4
10	97.3 $\pm$ 32.5	133.2 $\pm$ 27	154.9 $\pm$ 59.2	10 $\pm$ 6	26 $\pm$ 16.7	125.5 $\pm$ 27.7

Non-GLP RNA dose ranging study (higher doses) in Indian rhesus macaques

[00453] This study was designed to (a) evaluate the immunogenicity of VEE-MAG25mer srRNA at a dose of 300 µg as a homologous prime/boost or heterologous prime/boost in combination with ChAdV68.5WTnt.MAG25mer; (b) compare the immune responses of VEE-MAG25mer srRNA in lipid nanoparticles using LNP1 versus LNP2 at the 300µg dose; and (c) evaluate the kinetics of T-cell responses to VEE-MAG25mer srRNA and ChAdV68.5WTnt.MAG25mer immunizations.

[00454] The study arm was conducted in Mamu-A*01 Indian rhesus macaques to demonstrate immunogenicity. Vaccine immunogenicity in nonhuman primate species, such as Rhesus, is the best predictor of vaccine potency in humans. Furthermore, select antigens used in this study are only recognized in Rhesus macaques, specifically those with a Mamu-A*01 MHC class I haplotype. Mamu-A*01 Indian rhesus macaques were randomized to the different study arms (6 macaques per group) and administered an IM injection bilaterally with either ChAdV68.5-WTnt.MAG25mer or VEE-MAG25mer srRNA encoding model antigens that includes multiple Mamu-A*01 restricted antigens. The study arms were as described below.

[00455] PBMCs were collected prior to immunization and 4, 5, 6, 7, 8, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23 or 24 weeks after the initial immunization for immune monitoring for group 1 (heterologous prime/boost). PBMCs were collected prior to immunization and 4, 5, 7, 8, 10, 11, 12, 13, 14, or 15 weeks after the initial immunization for immune monitoring for groups 2 and 3 (homologous prime/boost).

[00456] **Table F:** Non-GLP immunogenicity study in Indian Rhesus Macaques

Group	Prime	Boost 1	Boost 2	Boost 3
1	ChAdV68.5WTnt.MAG25mer	VEE-MAG25mer srRNA -LNP2 (300 µg)	VEE-MAG25mer srRNA -LNP2 (300 µg)	VEE-MAG25mer srRNA -LNP2 (300 µg)
2	VEE-MAG25mer srRNA -LNP2 (300 µg)	VEE-MAG25mer srRNA -LNP2 (300 µg)	VEE-MAG25mer srRNA -LNP2 (300 µg)	
3	VEE-MAG25mer srRNA -LNP1 (300 µg)	VEE-MAG25mer srRNA -LNP1 (300 µg)	VEE-MAG25mer srRNA -LNP1 (300 µg)	

Results

[00457] Mamu-A*01 Indian rhesus macaques were immunized with ChAdV68.5-WTnt.MAG25mer. Antigen-specific cellular immune responses in peripheral blood mononuclear cells (PBMCs) were measured to six different Mamu-A*01 restricted epitopes prior to immunization and 4, 5, 6, 7, 8, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23 or 24 weeks after the initial immunization (FIG. 2 and Table G). Animals received boost immunizations with VEE-MAG25mer srRNA using the LNP2 formulation on weeks 4, 12, and 20. Combined antigen-specific immune responses of 1750, 4225, 1100, 2529, 3218, 1915, 1708, 1561, 5077, 4543, 4920, 5820, 3395, 2728, 1996, 1465, 4730, 2984, 2828, or 3043 SFCs per 10^6 PBMCs (six epitopes combined) were measured 4, 5, 6, 7, 8, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23 or 24 weeks after the initial immunization with ChAdV68.5WTnt.MAG25mer (FIG. 2). Immune responses measured 1 week after the second boost immunization (week 13) with VEE-MAG25mer srRNA were ~3-fold higher than that measured just before the boost immunization (week 12). Immune responses measured 1 week after the third boost immunization (week 21) with VEE-MAG25mer srRNA, were ~3-fold higher than that measured just before the boost immunization (week 20), similar to the response observed for the second boost.

[00458] Mamu-A*01 Indian rhesus macaques were also immunized with VEE-MAG25mer srRNA using two different LNP formulations (LNP1 and LNP2). Antigen-specific cellular immune responses in peripheral blood mononuclear cells (PBMCs) were measured to six different Mamu-A*01 restricted epitopes prior to immunization and 4, 5, 6, 7, 8, 10, 11, 12, 13, 14, or 15 weeks after the initial immunization (FIGS. 3 and 4, Tables H and I). Animals received boost immunizations with VEE-MAG25mer srRNA using the respective LNP1 or LNP2 formulation on weeks 4 and 12. Combined antigen-specific immune responses of 168, 204, 103, 126, 140, 145, 330, 203, and 162 SFCs per 10^6 PBMCs (six epitopes combined) were measured 4, 5, 7, 8, 10, 11, 13, 14, 15 weeks after the immunization with VEE-MAG25mer srRNA-LNP2 (FIG. 3). Combined antigen-specific immune responses of 189, 185, 349, 437, 492, 570, 233, 886, 369, and 381 SFCs per 10^6 PBMCs (six epitopes combined) were measured 4, 5, 7, 8, 10, 11, 12, 13, 14, 15 weeks after the immunization with VEE-MAG25mer srRNA-LNP1 (FIG. 4).

[00459] Table G: Mean spot forming cells (SFC) per 10^6 PBMCs for each epitope $\pm$ SEM for priming vaccination with ChAdV68.5WTnt.MAG25mer (Group 1)

Wk	Antigen					
	Env CL9	Env TL9	Gag CM9	Gag LW9	Pol SV9	Tat TL8
4	173 $\pm$ 41.6	373.5 $\pm$ 87.3	461.4 $\pm$ 74.2	38.4 $\pm$ 26.1	94.5 $\pm$ 26	609.2 $\pm$ 121.9
5	412.7 $\pm$ 138.4	987.8 $\pm$ 283.3	1064.4 $\pm$ 266.9	85.6 $\pm$ 31.2	367.2 $\pm$ 135.2	1306.8 $\pm$ 332.8
6	116.2 $\pm$ 41.2	231.1 $\pm$ 46.3	268.3 $\pm$ 90.7	86.1 $\pm$ 42	174.3 $\pm$ 61	223.9 $\pm$ 38.1
7	287.4 $\pm$ 148.7	588.9 $\pm$ 173.9	693.2 $\pm$ 224.8	92.1 $\pm$ 33.5	172.9 $\pm$ 55.6	694.6 $\pm$ 194.8
8	325.4 $\pm$ 126.6	735.8 $\pm$ 212	948.9 $\pm$ 274.5	211.3 $\pm$ 62.7	179.1 $\pm$ 50	817.3 $\pm$ 185.2
10	312 $\pm$ 129.7	543.2 $\pm$ 188.4	618.6 $\pm$ 221.7	-5.7 $\pm$ 4.1	136.5 $\pm$ 51.3	309.9 $\pm$ 85.6
11	248.5 $\pm$ 81.1	348.7 $\pm$ 129.8	581.1 $\pm$ 205.5	-3.1 $\pm$ 4.4	119 $\pm$ 51.2	413.7 $\pm$ 144.8
12	261.9 $\pm$ 68.2	329.9 $\pm$ 83	486.5 $\pm$ 118.6	-1.2 $\pm$ 5.1	132.8 $\pm$ 31.8	350.9 $\pm$ 69.3
13	389.3 $\pm$ 167.7	1615.8 $\pm$ 418.3	1244.3 $\pm$ 403.6	1.3 $\pm$ 8.1	522.5 $\pm$ 155	1303.3 $\pm$ 385.6
14	406.3 $\pm$ 121.6	1616 $\pm$ 491.7	1142.3 $\pm$ 247.2	6.6 $\pm$ 11.1	322.7 $\pm$ 94.1	1048.6 $\pm$ 215.6
15	446.8 $\pm$ 138.7	1700.8 $\pm$ 469.1	1306.3 $\pm$ 294.4	43 $\pm$ 24.5	421.2 $\pm$ 87.9	1001.5 $\pm$ 236.4
16	686.8 $\pm$ 268.8	1979.5 $\pm$ 541.7	1616.8 $\pm$ 411.8	2.4 $\pm$ 7.8	381.9 $\pm$ 116.4	1152.8 $\pm$ 352.7
17	375.8 $\pm$ 109.3	1378.6 $\pm$ 561.2	773.1 $\pm$ 210.3	-1.4 $\pm$ 4.3	177.6 $\pm$ 93.7	691.7 $\pm$ 245
18	255.9 $\pm$ 99.7	1538.4 $\pm$ 498.1	498.7 $\pm$ 152.3	-5.3 $\pm$ 3.3	26.2 $\pm$ 13.4	413.9 $\pm$ 164.8
19	133 $\pm$ 62.6	955.9 $\pm$ 456.8	491.1 $\pm$ 121.8	-5.7 $\pm$ 4.1	50.3 $\pm$ 25.4	371.2 $\pm$ 123.7
20	163.7 $\pm$ 55.8	641.7 $\pm$ 313.5	357.9 $\pm$ 91.1	2.6 $\pm$ 7.5	41.4 $\pm$ 24.2	257.8 $\pm$ 68.9
21	319.9 $\pm$ 160.5	2017.1 $\pm$ 419.9	1204.8 $\pm$ 335.2	-3.7 $\pm$ 5.1	268.1 $\pm$ 109.6	924.1 $\pm$ 301
22	244.7 $\pm$ 105.6	1370.9 $\pm$ 563.5	780.3 $\pm$ 390	-3.6 $\pm$ 5.1	118.2 $\pm$ 68.1	473.3 $\pm$ 249.3
23	176.7 $\pm$ 81.8	1263.7 $\pm$ 527.3	838.6 $\pm$ 367.9	-5.7 $\pm$ 4.1	73.6 $\pm$ 49	480.9 $\pm$ 163.9
24	236.5 $\pm$ 92	1324.7 $\pm$ 589.3	879.7 $\pm$ 321	-0.4 $\pm$ 5.7	104 $\pm$ 53.1	498 $\pm$ 135.8

[00460] Table H: Mean spot forming cells (SFC) per 10^6 PBMCs for each epitope $\pm$ SEM for priming vaccination with VEE-MAG25mer srRNA-LNP2 (300 μ g) (Group 2)

Wk	Antigen					
	Env CL9	Env TL9	Gag CM9	Gag LW9	Pol SV9	Tat TL8
4	46 $\pm$ 27.1	18.4 $\pm$ 6.8	58.3 $\pm$ 45.8	29.9 $\pm$ 20.8	4.9 $\pm$ 2.3	10.7 $\pm$ 4
5	85.4 $\pm$ 54	5.2 $\pm$ 5.8	52.4 $\pm$ 51.2	34.5 $\pm$ 35	11.8 $\pm$ 12.2	14.4 $\pm$ 7.9
7	18.6 $\pm$ 32.5	1.9 $\pm$ 1.7	59.4 $\pm$ 55.7	9.3 $\pm$ 10.7	3.3 $\pm$ 3	10.7 $\pm$ 6.1
8	36.6 $\pm$ 39.4	6.3 $\pm$ 3.9	48.7 $\pm$ 39.9	13.5 $\pm$ 8.8	3.8 $\pm$ 3.6	17.2 $\pm$ 9.7
10	69.1 $\pm$ 59.1	4.4 $\pm$ 1.9	39.3 $\pm$ 38	14.7 $\pm$ 10.8	4.4 $\pm$ 5.3	8.5 $\pm$ 5.3
11	43 $\pm$ 38.8	22.6 $\pm$ 21.1	30.2 $\pm$ 26.2	3.3 $\pm$ 2.2	5.8 $\pm$ 3.5	40.3 $\pm$ 25.5
13	120.4 $\pm$ 78.3	68.2 $\pm$ 43.9	54.2 $\pm$ 36.8	21.8 $\pm$ 7.4	17.7 $\pm$ 6.1	47.4 $\pm$ 27.3
14	76 $\pm$ 44.8	28 $\pm$ 19.5	65.9 $\pm$ 64.3	-0.3 $\pm$ 1.3	2.5 $\pm$ 2	31.1 $\pm$ 26.5
15	58.9 $\pm$ 41.4	19.5 $\pm$ 15.1	55.4 $\pm$ 51	2.5 $\pm$ 2	5.5 $\pm$ 3.6	20.1 $\pm$ 15.7

[00461] Table I: Mean spot forming cells (SFC) per 10^6 PBMCs for each epitope $\pm$ SEM for priming vaccination with VEE-MAG25mer srRNA-LNP1 (300 μ g) (Group 3)

Wk	Antigen					
	Env CL9	Env TL9	Gag CM9	Gag LW9	Pol SV9	Tat TL8
4	19.5 $\pm$ 8.7	13.3 $\pm$ 13.1	16.5 $\pm$ 15.3	10.5 $\pm$ 7.3	35.9 $\pm$ 24.8	92.9 $\pm$ 91.6
5	87.9 $\pm$ 43.9	12.7 $\pm$ 11.7	37.2 $\pm$ 31.9	21.1 $\pm$ 23.8	13.2 $\pm$ 13.7	12.6 $\pm$ 13.7
7	21.1 $\pm$ 13.3	48.8 $\pm$ 48.4	51.7 $\pm$ 39.5	9.1 $\pm$ 10.5	58.6 $\pm$ 55.8	159.4 $\pm$ 159
8	47.7 $\pm$ 21.7	66.4 $\pm$ 52.2	59.8 $\pm$ 57.4	49.4 $\pm$ 28	79.4 $\pm$ 63	133.8 $\pm$ 132.1
10	49 $\pm$ 30.2	42.2 $\pm$ 41.1	139.3 $\pm$ 139.3	51.6 $\pm$ 51.2	78.2 $\pm$ 75.8	131.7 $\pm$ 131.6
11	42 $\pm$ 26.8	20.9 $\pm$ 21.4	177.1 $\pm$ 162	-6.3 $\pm$ 4.3	104.3 $\pm$ 104.1	231.5 $\pm$ 230.1
12	40.2 $\pm$ 19	20.3 $\pm$ 11.9	42.2 $\pm$ 46.7	3.7 $\pm$ 6.7	57 $\pm$ 44.7	70 $\pm$ 69.2
13	81.2 $\pm$ 48.9	38.2 $\pm$ 37.6	259.4 $\pm$ 222.2	-4 $\pm$ 4.1	164.1 $\pm$ 159.3	347.3 $\pm$ 343.5
14	34.5 $\pm$ 31.8	5.3 $\pm$ 11.6	138.6 $\pm$ 137.3	-4.7 $\pm$ 5.2	52.3 $\pm$ 52.9	142.6 $\pm$ 142.6
15	49 $\pm$ 24	6.7 $\pm$ 9.8	167.1 $\pm$ 163.8	-6.4 $\pm$ 4.2	47.8 $\pm$ 42.3	116.6 $\pm$ 114.5

srRNA Dose Ranging Study

[00462] In one implementation of the present invention, an srRNA dose ranging study can be conducted in mamu A01 Indian rhesus macaques to identify which srRNA dose to progress to NHP immunogenicity studies. In one example, Mamu A01 Indian rhesus macaques can be administered with an srRNA vector encoding model antigens that includes multiple mamu A01 restricted epitopes by IM injection. In another example, an anti-CTLA-4 monoclonal antibody can be administered SC proximal to the site of IM vaccine injection to target the vaccine draining lymph node in one group of animals. PBMCs can be collected every 2 weeks after the initial vaccination for immune monitoring. The study arms are described in below (Table J).

Table J: Non-GLP RNA dose ranging study in Indian Rhesus Macaques

Group	Prime	Boost 1	Boost 2
1	srRNA-LNP (Low Dose)	srRNA-LNP (Low Dose)	srRNA-LNP (Low Dose)
2	srRNA-LNP (Mid Dose)	srRNA-LNP (Mid Dose)	srRNA-LNP (Mid Dose)
3	srRNA-LNP (High Dose)	srRNA-LNP (High Dose)	srRNA-LNP (High Dose)
4	srRNA-LNP (High Dose) + anti-CTLA-4	srRNA-LNP (High Dose) + anti-CTLA-4	srRNA-LNP (High Dose) + anti-CTLA-4

* Dose range of srRNA to be determined with the high dose ≤ 300 μ g.

Immunogenicity Study in Indian Rhesus Macaques

[00463] Vaccine studies were conducted in mamu A01 Indian rhesus macaques (NHPs) to demonstrate immunogenicity using the antigen vectors. Fig. 5 illustrates the vaccination strategy. Three groups of NHPs were immunized with ChAdV68.5-WTnt.MAG25mer and either with the checkpoint inhibitor anti-CTLA-4 antibody Ipilimumab (Groups 5 & 6) or without the checkpoint inhibitor (Group 4). The antibody was administered either intravenously (group 5) or subcutaneously (group 6). Triangles indicate chAd68 vaccination (1×10^{12} vp/animal) at weeks 0 & 32. Circles represent alphavirus vaccination at weeks 0, 4, 12, 20, 28 and 32.

[00464] The time course of CD8⁺ anti-epitope responses in the immunized NHPs are presented for chAd-MAG immunization alone (Fig. 6 and Table K), chAd-MAG immunization with the checkpoint inhibitor delivered IV (Fig. 7 and Table L), and chAd-MAG immunization

with the checkpoint inhibitor delivered SC (Fig. 8 and Table M). The results demonstrate chAd68 vectors efficiently primed CD8⁺ responses in primates, alphavirus vectors efficiently boosted the chAd68 vaccine priming response, checkpoint inhibitor whether delivered IV or SC amplified both priming and boosting responses, and chAd vectors readministered post vaccination to effectively boosted the immune responses.

Table K: CD8⁺ anti-epitope responses in Rhesus Macaques dosed with chAd-MAG (Group 4). Mean SFC/1e6 splenocytes +/- the standard error is shown

Wk	Antigen					
	Env CL9	Env TL9	Gag CM9	Gag LW9	Pol SV9	Tat TL8
4	173±41.6	373.5±87.3	461.4±74.2	38.4±26.1	94.5±26	609.2±121.9
5	412.7±138.4	987.8±283.3	1064.4±266.9	85.6±31.2	367.2±135.2	1306.8±332.8
6	116.2±41.2	231.1±46.3	268.3±90.7	86.1±42	174.3±61	223.9±38.1
7	287.4±148.7	588.9±173.9	693.2±224.8	92.1±33.5	172.9±55.6	694.6±194.8
8	325.4±126.6	735.8±212	948.9±274.5	211.3±62.7	179.1±50	817.3±185.2
10	312±129.7	543.2±188.4	618.6±221.7	-5.7±4.1	136.5±51.3	309.9±85.6
11	248.5±81.1	348.7±129.8	581.1±205.5	-3.1±4.4	119±51.2	413.7±144.8
12	261.9±68.2	329.9±83	486.5±118.6	-1.2±5.1	132.8±31.8	350.9±69.3
13	389.3±167.7	1615.8±418.3	1244.3±403.6	1.3±8.1	522.5±155	1303.3±385.6
14	406.3±121.6	1616±491.7	1142.3±247.2	6.6±11.1	322.7±94.1	1048.6±215.6
15	446.8±138.7	1700.8±469.1	1306.3±294.4	43±24.5	421.2±87.9	1001.5±236.4
16	686.8±268.8	1979.5±541.7	1616.8±411.8	2.4±7.8	381.9±116.4	1152.8±352.7
17	375.8±109.3	1378.6±561.2	773.1±210.3	-1.4±4.3	177.6±93.7	691.7±245
18	255.9±99.7	1538.4±498.1	498.7±152.3	-5.3±3.3	26.2±13.4	413.9±164.8
19	133±62.6	955.9±456.8	491.1±121.8	-5.7±4.1	50.3±25.4	371.2±123.7
20	163.7±55.8	641.7±313.5	357.9±91.1	2.6±7.5	41.4±24.2	257.8±68.9
21	319.9±160.5	2017.1±419.9	1204.8±335.2	-3.7±5.1	268.1±109.6	924.1±301
22	244.7±105.6	1370.9±563.5	780.3±390	-3.6±5.1	118.2±68.1	473.3±249.3
23	176.7±81.8	1263.7±527.3	838.6±367.9	-5.7±4.1	73.6±49	480.9±163.9
24	236.5±92	1324.7±589.3	879.7±321	-0.4±5.7	104±53.1	498±135.8
25	136.4±52.6	1207.1±501.6	924±358.5	6.2±10.5	74.1±34.4	484.6±116.7
26	278.2±114.4	1645±661.7	1170.2±469.9	-2.9±5.7	80.6±55.8	784.4±214.1
27	159±56.8	961.7±547.1	783.6±366.4	-5±4.3	63.6±27.5	402.9±123.4
28	189.6±75.7	1073.1±508.8	668.3±312.5	-5.7±4.1	80.3±38.3	386.4±122
29	155.3±69.1	1102.9±606.1	632.9±235	34.5±24.2	80±35.5	422.5±122.9
30	160.2±59.9	859±440.9	455±209.1	-3±5.3	60.5±28.4	302.7±123.2
31	122.2±49.7	771.1±392.7	582.2±233.5	-5.7±4.1	55.1±27.3	295.2±68.3
32	119.3±28.3	619.4±189.7	566±222.1	-3.7±5.1	21.9±4.5	320.5±76.4
33	380.5±122	1636.1±391.4	1056.2±205.7	-5.7±4.1	154.5±38.5	988.4±287.7
34	1410.8±505.4	972.4±301.5	319.6±89.6	-4.8±4.2	141.1±49.8	1375.5±296.7
37	130.8±29.2	500±156.9	424.9±148.9	-3.5±4.7	77.7±24.6	207.1±42.4
38	167.7±54.8	1390.8±504.7	830.4±329.1	-5.5±4.1	111.8±43.2	516±121.7

Table L: CD8⁺ anti-epitope responses in Rhesus Macaques dosed with chAd-MAG plus anti-CTLA4 antibody (Ipilimumab) delivered IV.(Group 5). Mean SFC/1e6 splenocytes +/- the standard error is shown

Wk	Antigen					
	Env CL9	Env TL9	Gag CM9	Gag LW9	Pol SV9	Tat TL8
4	1848.1±432.2	1295.7±479.7	1709.8±416.9	513.7±219.8	838.5±221.1	2514.6±246.5
5	1844.1±410.2	2367.5±334.4	1983.1±370.7	732.1±249.4	1429.7±275.3	2517.7±286.5
6	822.4±216.7	1131.2±194.7	796.8±185.8	226.8±70	802.2±101.4	913.5±222.7
7	1147.2±332.9	1066±311.2	1149.8±467.3	267.4±162.6	621.5±283.2	1552.2±395.1
8	1192.7±188.8	1461.5±237.7	1566.9±310.5	522.5±118.6	662.3±142.4	1706±216.7
10	1249±220.3	1170.6±227.7	1297.3±264.7	-0.3±4.4	551.8±90.5	1425.3±142.6
11	934.2±221.7	808±191.3	1003.1±293.4	1.9±4.3	364.2±76.6	1270.8±191.6
12	1106.2±216.6	896.7±190.7	1020.1±243.3	1.3±3.9	436.6±90	1222±155.4
13	2023.8±556.3	3696.7±1.7	2248.5±436.4	-4.5±3.5	2614±406.1	3700±0
14	1278.7±240	2639.5±387	1654.6±381.1	-6±2.1	988.8±197.9	2288.3±298.7
15	1458.9±281.8	2932.5±488.7	1893.4±499	74.6±15.6	1657.8±508.9	2709.1±428.7
16	1556.8±243	2143.8±295.2	2082.8±234.2	-5.8±2.5	1014.6±161.4	2063.7±86.7
17	1527±495.1	2213±677.1	1767.7±391.8	15.1±5.9	633.8±133.9	2890.8±433.9
18	1068.2±279.9	1940.9±204.1	1114.1±216.1	-5.8±2.5	396.6±77.6	1659.4±171.7
19	760.7±362.2	1099.5±438.4	802.7±192.5	-2.4±3.3	262.2±62.2	1118.6±224.2
20	696.3±138.2	954.9±198	765.1±248.4	-1.4±4.4	279.6±89.3	1139±204.5
21	1201.4±327.9	3096±1.9	1901±412.1	-5.8±2.5	1676.3±311.5	2809.3±195.8
22	1442.5±508.3	2944.7±438.6	1528.4±349.6	2.8±5.1	940.7±160.5	2306.3±218.6
23	1400.4±502.2	2757.1±452.9	1604.2±450.1	-5.1±2.3	708.1±162.6	2100.4±362.9
24	1351±585.1	2264.5±496	1080.6±253.8	0.3±6.5	444.2±126.4	1823.7±306.5
25	1211.5±505.2	2160.4±581.8	970.8±235.9	2.5±3.8	450.4±126.9	1626.2±261.3
26	1443±492.5	2485±588	1252.5±326.4	-0.2±6	360.2±92.3	2081.9±331.1
27	896.2±413.3	1686±559.5	751±192.1	-3.7±2.5	247.4±82.8	1364.1±232
28	1147.8±456.9	1912.1±417.1	930.3±211.4	-5.8±2.5	423.9±79.6	1649.3±315
29	1038.5±431.9	1915.2±626.1	786.8±205.9	23.5±8.3	462.8±64	1441.5±249.7
30	730.5±259.3	1078.6±211.5	699.1±156.2	-4.4±2.7	234.4±43.9	1160.6±112.6
31	750.4±328.3	1431±549.9	650.6±141.1	-5.2±3	243.4±56.4	868.9±142.8
32	581.4±227.7	1326.6±505.2	573.3±138	-3.2±4.2	160.8±49.2	936.4±110.4
33	2198.4±403.8	3093.4±123.3	2391.8±378.4	7.1±8.5	1598.1±343.1	2827.5±289.5
34	2654.3±337	2709.9±204.3	1297.5±291.4	0.4±4.2	1091.8±242.9	1924±245.7
37	846.8±301.7	1706.9±196	973.6±149.3	50.5±45.2	777.3±140.2	1478.8±94.3

Table M: CD8⁺ anti-epitope responses in Rhesus Macaques dosed with chAd-MAG plus anti-CTLA4 antibody (Ipilimumab) delivered SC (Group 6). Mean SFC/1e6 splenocytes +/- the standard error is shown

Wk	Antigen					
	Env CL9	Env TL9	Gag CM9	Gag LW9	Pol SV9	Tat TL8
4	598.3±157.4	923.7±306.8	1075.6±171.8	180.5±74.1	752.3±245.8	1955.3±444.4
5	842.2±188.5	1703.7±514.2	1595.8±348.4	352.7±92.3	1598.9±416.8	2163.7±522.1
6	396.4±45.3	728.3±232.7	503.8±151.9	282±69	463.1±135.7	555.2±191.5
7	584.2±177	838.3±254.9	1013.9±349.4	173.6±64.3	507.4±165.7	1222.8±368
8	642.9±134	1128.6±240.6	1259.1±163.8	366.1±72.8	726.7±220.9	1695.6±359.4
10	660.4±211.4	746.9±222.7	944.8±210.2	-1.2±1.9	523.4±230.7	787.3±308.3
11	571.2±162	609.4±194.3	937.9±186.5	-8.9±2.3	511.6±229.6	1033.3±315.7
12	485.3±123.7	489.4±142.7	919.3±214.1	-8.9±2.3	341.6±139.4	1394.7±432.1
13	986.9±154.5	2811.9±411.3	1687.7±344.3	-4.1±5.1	1368.5±294.2	2751±501.9
14	945.9±251.4	2027.7±492.8	1386.7±326.7	-5.7±2.8	708.9±277.1	1588.2±440.1
15	1075.2±322.4	2386±580.7	1606.3±368.1	-5.4±3.2	763.3±248.8	1896.5±507.8
16	1171.8±341.6	2255.1±439.6	1672.2±342.3	-7.8±2.4	1031.6±228.8	1896.4±419.9
17	1118.2±415.4	2156.3±476	1345.3±377.7	-1.1±6.7	573.7±118.8	1614.4±382.3
18	861.3±313.8	2668.2±366.8	1157.2±259.6	-8.9±2.3	481.2±164	1725.8±511.4
19	719.2±294.2	1447.2±285	968±294.5	-2.2±4.6	395.6±106.1	1199.6±289.2
20	651.6±184	1189.8±242.8	947.4±249.8	-8.9±2.3	355±106.3	1234.7±361.7
21	810.3±301.9	2576.2±283.7	1334±363.1	-8.9±2.3	892.2±305	1904.4±448.1
22	775±196.4	2949±409.7	1421.8±309.7	38±27.8	577±144.2	2330.6±572.3
23	584.9±240.2	1977.9±361.4	1209.8±405.1	-7.3±3.2	273.7±93.3	1430.6±363.9
24	485.4±194.4	1819.8±325.5	837.2±261.4	-3.4±4.1	234.4±71.1	943.9±243.3
25	452.3±175	2072±405.7	957.1±293.1	-8.9±2.3	163±43.2	1341.2±394.7
26	517.9±179.1	2616±567.5	1126.6±289	-8.3±2.3	199.9±89.2	1615.7±385.6
27	592.8±171.7	1838.3±372.4	749.3±170.4	-7.3±2.5	325.5±98.7	1110.7±308.8
28	793±228.5	1795.4±332.3	1068.7±210.3	2.5±4.1	553.1±144.3	1480.8±357.1
29	661.8±199.9	2140.6±599.3	1202.7±292.2	-8.7±2.8	558.9±279.2	1424.2±408.6
30	529.1±163.3	1528.2±249.8	840.5±218.3	-8.9±2.3	357.7±149.4	1029.3±335
31	464.8±152.9	1332.2±322.7	726.3±194.3	-8.9±2.3	354.3±158.6	884.4±282
32	612.9±175.3	1584.2±390.2	1058.3±219.8	-8.7±2.8	364.6±149.8	1388.8±467.3
33	1600.2±416.7	2597.4±367.9	2086.4±414.8	-6.3±3.3	893.8±266	2490.6±416.4
34	2814.6±376.2	2713.6±380.8	1888.8±499.4	-7.5±3.1	1288.9±438.9	2428.1±458.9
37	1245.9±471.7	1877.7±291.2	1606.6±441.9	14.2±13	1227.5±348.1	1260.7±342

Memory Phenotyping in Indian Rhesus Macaques

[00465] Rhesus macaque were immunized with ChAdV68.5WTnt.MAG25mer /VEE-MAG25mer srRNA heterologous prime/boost regimen with or without anti-CTLA4, and boosted again with ChAdV68.5WTnt.MAG25mer. Groups were assessed 11 months after the final ChAdV68 administration (study month 18). by ELISpot was performed as described. Fig. 9 and Table N shows cellular responses to six different Mamu-A*01 restricted epitopes as measured by ELISpot both pre-immunization (left panel) and after 18 months (right panel). The detection of responses to the restricted epitopes demonstrates antigen-specific memory responses were generated by ChAdV68/samRNA vaccine protocol.

[00466] To assess memory, CD8⁺ T-cells recognizing 4 different rhesus macaque Mamu-A*01 class I epitopes encoded in the vaccines were monitored using dual-color Mamu-A*01 tetramer labeling, with each antigen being represented by a unique double positive combination, and allowed the identification of all 4 antigen-specific populations within a single sample. Memory cell phenotyping was performed by co-staining with the cell surface markers CD45RA and CCR7. Fig. 10 and Table O shows the results of the combinatorial tetramer staining and CD45RA/CCR7 co-staining for memory T-cells recognizing four different Mamu-A*01 restricted epitopes. The T cell phenotypes were also assessed by flow cytometry. Fig. 11 shows the distribution of memory cell types within the sum of the four Mamu-A*01 tetramer+ CD8⁺ T-cell populations at study month 18. Memory cells were characterized as follows: CD45RA+CCR7+=naïve, CD45RA+CCR7-=effector (Teff), CD45RA-CCR7+=central memory (Tcm), CD45RA-CCR7-=effector memory (Tem). Collectively, the results demonstrate that memory responses were detected at least one year following the last boost indicating long lasting immunity, including effector, central memory, and effector memory populations.

Table N Mean spot forming cells (SFC) per 10⁶ PBMCs for each animal at both pre-prime and memory assessment time points (18 months).

Animal	Pre-prime baseline						18 months					
	Tat TL8	Gag CM9	Env TL9	Env CL9	Gag LW9	Pol SV9	Tat TL8	Gag CM9	Env TL9	Env CL9	Gag LW9	Pol SV9
1	1.7	0.0	0.0	5.0	0.0	13.7	683.0	499.2	1100.3	217.5	47.7	205.3
2	0.0	0.0	0.0	0.2	0.1	0.0	773.4	ND	1500.0	509.3	134.5	242.5
3	0.0	0.0	6.7	6.8	10.2	3.3	746.3	167.5	494.1	402.8	140.6	376.0
4	0.0	0.0	0.0	0.0	0.0	0.0	47.6	1023.9	85.1	44.2	44.2	47.6
5	15.3	6.7	18.6	15.6	5.2	12.1	842.4	467.7	1500.0	805.9	527.8	201.8
6	3.1	0.0	0.0	15.5	6.9	5.3	224.3	720.3	849.0	296.9	32.4	121.9

ND=not determined due to technical exclusion

Table O Percent Mamu-A*01 tetramer positive out of live CD8⁺ cells

Animal	Tat TL8	Gag CM9	Env TL9	Env CL9
1	0.42	0.11	0.19	0.013
2	0.36	0.048	0.033	0.00834
3	0.97	0.051	0.35	0.048
4	0.46	0.083	0.17	0.028
5	0.77	0.45	0.14	0.2
6	0.71	0.16	0.17	0.04

XVII. Influenza Studies

[00467] Various dosing protocols using ChAdV68 and self-replicating RNA (srRNA) were evaluated for influenza vaccines.

Materials and Methods

Influenza ChAdV68 and SAM Vectors

[00468] ChAdV68 and SAM influenza vectors expressing influenza hemagglutinins (HA) from the following influenza A strains were made: H1N1 California7/09 and H7N9 Anhui1/13. The HA nucleotide sequences were codon optimized for optimized mammalian expression of the HA protein (*see Sequences*). The constructs were cloned by Gibson assembly into ChAdV68 and SAM plasmids. The ChAdV68 vector backbone used was the following: AC_000011.1 with E1 (nt 577 to 3403) and E3 (nt 27,125- 31,825) sequences deleted; corresponding ATCC VR-594 nucleotides substituted at five positions; antigen under the control of the CMV promoter/enhancer inserted in place of deleted E1. ChAdV68 plasmids were PacI digested and transfected into 293F cells to produce ChAd-HA viral vectors. ChAd production was performed at the 400 mL scale in 293F suspension cells. Virus was purified by two rounds of discontinuous CsCl density gradients and the virus was purified and dialyzed into ARM buffer (10 mM Tris pH8.0, 25 mM NaCl, 2.5% Glycerol). Vector was diluted into ARM buffer to the target dose for administration to animals. The SAM vector backbone used was the following: Venezuelan Equine Encephalitis (VEE) (Kinney, 1986, Virology 152: 400-413) based self-amplifying RNA (SAM) vector with sequences encoding the structural proteins of VEE located 3' of the 26S sub-genomic promoter were deleted (VEE sequences 7544 to 11,175 deleted; numbering based on Kinney *et al* 1986; SEQ ID NO:6) and replaced by antigen sequences. A high-quality full-length SAM encoding HA of H1N1 California7/09 or H7N9 Anhui1/13 strain was synthesized and co-transcriptionally capped *in vitro* and purified through silica columns. SAM was then encapsulated in lipid nanoparticles (LNP) and stored in a storage buffer.

In Vivo Evaluation in Balb/C mice

[00469] For humoral assessment, ChAd-CA709-HA and ChAd-Anhui1/13-HA vectors were dosed at either 5e10 VP or 1e9 VP/animal. Sera was harvested pre-injection and at day 28 post prime and at day 56, 4 weeks post SAM boost. Limiting dilutions of sera was analyzed for the

ability to prevent hemagglutination of red blood cells and a hemagglutination inhibition (HAI) titer determined.

[00470] For T cell response assessment, Balb/c mice were immunized with 5e10 VP ChAd-CA709-HA or 10 ug SAM-CA709-HA and antigen specific T cell immune response were evaluated. At 2-weeks post prime, splenocytes were isolated and T cell response was assessed by overnight IFN γ ELISpot following stimulation with six minipools of overlapping peptides (15 AA long, 11 AA overlap) spanning the HA protein of influenza A CA/7/09.

[00471] Mice were also immunized in a ChAdV68-prime and SAM-boost regimen where mice were immunized with ChAd-HA (5e10 vp) and then boosted with SAM-HA at 8-weeks, with T cell responses assessed by IFN γ ELISpot at 2-weeks post boost (10 weeks post prime). T cell polyfunctionality was also assessed in splenocytes at 2-weeks post boost by intracellular cytokine staining (ICS) to identify T cells expressing IFN γ , TNF α or IL-2 post prime/boost with chAd & SAM-HA.

Results

[00472] Influenza vaccines were assessed *in vivo*.

[00473] Shown in FIG. 12 and quantified in Table P are HAI titers for influenza A H1N1 strain CA/7/09 pre and post immunization. Mice administered with the high dose of 5e10 vp/animal (FIG. 12A) demonstrated 100% of animals seroconverted and had high titers >300 HAI and all animals had HAI titers > 40 which is predicted to be protective. At the low dose (FIG. 12B), 5/6 animals had HAI titers >40 and these titers were greater than 4X the pre-samples indicating successful seroconversion that would likely be protective *in vivo*. Shown in FIG. 13 and quantified in Table Q are HAI titers for influenza A H7N9 Anhui 1/13 strain pre and post immunization (FIG. 13A high dose; FIG. 13B low dose). A similar, but generally lower, response was observed against the H7N9 strain with all animals seroconverting at the high dose of 5e10 vp. The results demonstrate ChAd-HA priming induced strong humoral immune responses against H1N1 and H7N9 influenza strains.

[00474] Shown in FIG. 14 and quantified in Table R are anti-HA T cell responses for influenza A H1N1 strain CA/7/09 with either a ChAdV68-prime (FIG. 14A), SAM-prime (FIG. 14B), or ChAdV68-prime and SAM-boost (FIG. 14C). ChAdV68-prime produced a more potent T-cell response across peptide pools as measured by interferon gamma ELISpot. SAM -prime produced a more variable, and generally reduced, T-cell response across peptide pools as measured by interferon gamma ELISpot. ChAdV68-prime and SAM-boost showed

significantly increased T cell responses compared to post prime responses with either vector. Shown in FIG. 15 and quantified in Table S are polyfunctional T cell responses for influenza A H1N1 strain CA/7/09 with a ChAdV68-prime and SAM-boost for CD8 T cells (FIG. 15A) and CD4 T cells (FIG. 15B). The T-cell response demonstrated a pronounced CD8 T cell bias. The results demonstrate the vaccine regimens induced strong, polyfunctional T cell responses against H1N1 influenza strains, in particular with ChAd-HA priming as well as subsequent boosting with a SAM-based vaccine.

Table P – H1N1 HAI titers pre and post immunization (“1-X” 5e10 VP; “2-X” 1e9 VP)

Mouse #	GMT Pre vaccination	GMT Post vaccination
1-1	5	320
1-2	5	320
1-3	5	640
1-4	5	640
1-5	5	320
1-6	5	320
2-1	5	5
2-2	5	160
2-3	5	80
2-4	5	40
2-5	5	160
2-6	5	40

Table Q – H7N9 HAI titers pre and post immunization (“3-X” 5e10 VP; “4-X” 1e9 VP)

Mouse #	GMT Pre vaccination	GMT Post vaccination
3-1	5	56.57
3-2	5	80
3-3	5	56.57
3-4	5	226.27
3-5	5	320
3-6	5	160
4-1	5	22.28
4-2	5	56.57
4-3	5	10
4-4	5	20
4-5	5	40
4-6	5	80

Table R – H1N1 T Cell Responses (interferon gamma ELISpot)

		Minipool								
Treatment	Animal ID	1	2	3	4	5	6	SUM	Mean	SE
ChAd-HA 5e10 vp	7	34	966	638	7	4116	2304	8065	10630	1732
	8	307	5526	1607	169	4790	5913	18312		
	9	81	1538	745	38	4926	2600	9927		
	10	59	1832	457	3	2463	1873	6687		
	11	59	3855	1307	22	5630	1556	12430		
	12	145	6100 ^a	2504	98	6100 ^a	5613	8360		
SAM-HA 10 ug	21	487	4240	1373	374	5366	4515	16355	6313	2467
	22	354	3649	1451	231	2213	3101	11000		
	23	49	1707	299	0	1332	476	3863		
	24	138	1519	504	30	176	1339	3706		
	25	104	648	217	0	87	666	1721		
	26	112	400	213	30	215	261	1232		
ChAd-HA + SAM-HA prime/boost	1	696	11991	2356	525	25000 ^a	4441	20009	45546	5690
	2	247	15232	3041	454	21429	14969	55371		
	3	276	16216	2078	465	23563	10576	53174		
	4	489	19196	2849	347	20170	9534	52584		
	5	292	12085	2109	572	20931	2514	38503		
	6	162	21544	2986	233	17148	11561	53635		

Table S – H1N1 Polyfunctional T Cell Responses (Cytokine ICS)

	% of CD8+			% of CD4+		
	IFNg	TNFa	IL2	IFNg	TNFa	IL2
	32.25	23.25	0.59	0.94	1.02	0.47
	25.67	16.70	0.47	0.90	1.01	0.44
	25.98	17.48	0.77	0.68	1.04	0.42
	31.97	22.10	0.55	1.38	1.43	0.61
	26.27	18.08	0.88	0.77	0.91	0.40
	40.89	26.66	1.02	0.99	0.81	0.35
Mean	30.51	20.71	0.71	0.94	1.04	0.45
SE	2.41	1.60	0.09	0.10	0.09	0.04

Cytokine responses are from individual animals and is the sum of the response to two overlapping peptide pools.

CERTAIN SEQUENCES

[00475] Vectors, cassettes, and antibodies referred to herein are described below and referred to by SEQ ID NO.

Tremelimumab VL (SEQ ID NO:16)
Tremelimumab VH (SEQ ID NO:17)
Tremelimumab VH CDR1 (SEQ ID NO:18)
Tremelimumab VH CDR2 (SEQ ID NO:19)
Tremelimumab VH CDR3 (SEQ ID NO:20)
Tremelimumab VL CDR1 (SEQ ID NO:21)
Tremelimumab VL CDR2 (SEQ ID NO:22)
Tremelimumab VL CDR3 (SEQ ID NO:23)
Durvalumab (MEDI4736) VL (SEQ ID NO:24)
MEDI4736 VH (SEQ ID NO:25)
MEDI4736 VH CDR1 (SEQ ID NO:26)
MEDI4736 VH CDR2 (SEQ ID NO:27)
MEDI4736 VH CDR3 (SEQ ID NO:28)
MEDI4736 VL CDR1 (SEQ ID NO:29)
MEDI4736 VL CDR2 (SEQ ID NO:30)
MEDI4736 VL CDR3 (SEQ ID NO:31)
UbA76-25merPDIT nucleotide (SEQ ID NO:32)
UbA76-25merPDIT polypeptide (SEQ ID NO:33)
MAG-25merPDIT nucleotide (SEQ ID NO:34)
MAG-25merPDIT polypeptide (SEQ ID NO:35)
Ub7625merPDIT NoSFL nucleotide (SEQ ID NO:36)
Ub7625merPDIT NoSFL polypeptide (SEQ ID NO:37)
ChAdV68.5WTnt.MAG25mer (SEQ ID NO:2); AC_000011.1 with E1 (nt 577 to 3403) and E3 (nt 27,125- 31,825) sequences deleted; corresponding ATCC VR-594 nucleotides substituted at five positions; model neoantigen cassette under the control of the CMV promoter/enhancer inserted in place of deleted E1; SV40 polyA 3' of cassette
Venezuelan equine encephalitis virus [VEE] (SEQ ID NO:3) GenBank: L01442.2
VEE-MAG25mer (SEQ ID NO:4); contains MAG-25merPDIT nucleotide (bases 30-1755)
Venezuelan equine encephalitis virus strain TC-83 [TC-83] (SEQ ID NO:5) GenBank: L01443.1
VEE Delivery Vector (SEQ ID NO:6); VEE genome with nucleotides 7544-11175 deleted [alphavirus structural proteins removed]
TC-83 Delivery Vector (SEQ ID NO:7); TC-83 genome with nucleotides 7544-11175 deleted [alphavirus structural proteins removed]
VEE Production Vector (SEQ ID NO:8); VEE genome with nucleotides 7544-11175 deleted, plus 5' T7-promoter, plus 3' restriction sites
TC-83 Production Vector (SEQ ID NO:9); TC-83 genome with nucleotides 7544-11175 deleted, plus 5' T7-promoter, plus 3' restriction sites
VEE-UbAAY (SEQ ID NO:14); VEE delivery vector with MHC class I mouse tumor epitopes SIINFELK and AH1-A5 inserted
VEE-Luciferase (SEQ ID NO:15); VEE delivery vector with luciferase gene inserted at 7545
ubiquitin (SEQ ID NO:38)>UbG76 0-228
Ubiquitin A76 (SEQ ID NO:39)>UbA76 0-228
HLA-A2 (MHC class I) signal peptide (SEQ ID NO:40)>MHC SignalPep 0-78
HLA-A2 (MHC class I) Trans Membrane domain (SEQ ID NO:41)>HLA A2 TM Domain 0-201
IgK Leader Seq (SEQ ID NO:42)>IgK Leader Seq 0-60
Human DC-Lamp (SEQ ID NO:43)>HumanDCLAMP 0-3178
Mouse LAMP1 (SEQ ID NO:44)>MouseLamp1 0-1858

Human Lamp1 cDNA (SEQ ID NO:45)>Human Lamp1 0-2339
Tetanus toxoid nucleic acid sequence (SEQ ID NO:46)
Tetanus toxoid amino acid sequence (SEQ ID NO:47)
PADRE nucleotide sequence (SEQ ID NO:48)
PADRE amino acid sequence (SEQ ID NO:49)
WPRES (SEQ ID NO:50)>WPRES 0-593
IRES (SEQ ID NO:51)>eGFP IRES SEAP Insert 1746-2335
GFP (SEQ ID NO:52)
SEAP (SEQ ID NO:53)
Firefly Luciferase (SEQ ID NO:54)
FMDV 2A (SEQ ID NO:55)
Ipilimumab Heavy Chain (SEQ ID NO: 29365)
QVQLVESGGGVVQPGRSLRLSCAASGFTFSSTMHVVRQAPGKGLEWVTFISYDGNKKYYADSVKGRFTISRDN SKNTLYLQMNSLRAEDTAIYYCARTGWLGPFDYWGQGLTVTVSSASTKGPSVFPLAPSSKSTSGGTAALGCLVK DYFPEPVTVSWNSGALTSGVHTFPAVLQSSGLYSLSVTVTPSSSLGTQTYICNVNHKPSNTKVDKRVKPKSCD KTHHTCPPCPAPELLGGPSVFLFPPKPKDITLMISRTPEVTCVVDVSHEDPEVKFNWYVDGVEVHNAKTKPREEQ YNSTYRVVSVLTVHLQDNLNGKEYKCKVSNKALPAPIEKTISKAKGQPREPQVYTLPPSRDELTKNQVSLTCLV KGFYPDSIAVEWESNGQPENNYKTTTPVLDSDGSFFLYSKLTVDKSRWQQGNVVFSCSVMEALHNHYTQKSLSL SPGK
Ipilimumab Light Chain (SEQ ID NO: 29366)
EIVLTQSPGTLSSLSPGERATLSCRASQSVGSSYLAWYQQKPGQAPRLLIYGAFSRATGIPDRFSGSGSGTDFTL TISRLEPEDFAVYYCQQYGSSPWTFGQGTKVEIKRTVAAPSVFIFPPSDEQLKSGTASVVCCLNNFYPREAKVQ WKVDNALQSGNSQESVTEQDSKSTYLSSTLTLSKADYEKHKVYACEVTHQGLSSPVTKSFNRGEC
Nivolumab Heavy Chain (SEQ ID NO: 29367)
QVQLVESGGGVVQPGRSLRLDCKASGITFSNSGMHVVRQAPGKGLEWVAVIWDGSKRYADSVKGRFTISRDN SKNTLFLQMNSLRAEDTAVYYCATNDDYWGQGLTVTVSSASTKGPSVFPLAPCSRSTSESTAALGCLVKDYFPE PVTVSWNSGALTSGVHTFPAVLQSSGLYSLSVTVTPSSSLGTQTYTCNVDHKPSNTKVDKRVESKYGPCCPPC PAPEFLGGPSVFLFPPKPKDITLMISRTPEVTCVVDVSDQEDPEVQFNWYVDGVEVHNAKTKPREEQFNSTYRVV SVLTVHLQDNLNGKEYKCKVSNKGLPSSIEKTISKAKGQPREPQVYTLPPSQEEMTKNQVSLTCLVKGFYPDSI AVEWESNGQPENNYKTTTPVLDSDGSFFLYSRLTVDKSRWQEGNVVFSCSVMEALHNHYTQKSLSLSLGK
Nivolumab Light Chain (SEQ ID NO: 29368)
EIVLTQSPATLSLSPGERATLSCRASQSVSSYLAWYQQKPGQAPRLLIYDASNRATGIPARFSGSGSGTDFTLT ISSLEPEDFAVYYCQQSSNWPRTFGQGTKVEIKRTVAAPSVFIFPPSDEQLKSGTASVVCCLNNFYPREAKVQW KVDNALQSGNSQESVTEQDSKSTYLSSTLTLSKADYEKHKVYACEVTHQGLSSPVTKSFNRGEC
Influenza A H1N1 CA/7/09 HA
ATGAAAGCTATCCTTGTAGTCCTTCTCTATACTTTTGCCACAGCTAACGCGGATACGCTGTGCATCGGGTATCA CGCAAATAACTCCACCGATACGGTGGACACGGTGCTTGAGAAGAATGTAAGTGAAGTCAATCCCGTGAAGTTC TGGAGGACAAACACAACGGGAAGTTGTGCAAGCTCAGGGGGGTGCGCCGTTGCATTTGGGAAAATGTAATATC GCTGGATGGATCTTGGGTAATCCCAGTGCAGAACGCTCAGCACCAGCAGAGCTGGAGCTACATTGTGGAAAC TCCCTCAAGCGATAACGGGACCTGTTACCCAGGGGATTTTCATCGACTATGAGGAATTGCGGGAACAGTTGAGTT CTGTGCTTTCATTTGAACGATTTGAAATTTTCCCAAGACCAGTTCTTGGCCTAATCACGACTCTAACAAGGGT GTTACGGCAGCATGCCCGCATGCCGGAGCAAAGAGTTTCTACAAAATTTGATCTGGTTGGTGAAGAAGGGCAA CTCATACCTTAAGCTCAGTAAGTCTTATATCAATGACAAAGGAAAGAAAGTACTGGTTTTGTGGGGAATCCACC ATCCGTCCACATCTGCAGACCAACAGTCACTTACCAAGACGCGGATGCCTACGTTTTGTGGGAAGCTCAAGA TATTCACAAAATTCAGCCTGAGATTGCTATTTCGCCAAAGGTCGCGACCAAGAAGGCAGGATGAATTACTA CTGGACCTTGGTCGAGCCTGGTGATAAGATTACATTTGAAGCCACCGGTAACCTTGTGTCCCGAGGTACGCCT TCGCGATGGAGCGGAATGCAGGGTCAGGGATTATTATATCAGACACCCAGTACACGACTGCAACACAAGTTGT CAGACCCCTAAGGGTGCCATCAATACATCCCTGCCGTTTTCAGAAATATCCATCCGATCACTATAGGCAAGTGTCC AAAATATGTGAAGAGCACGAAGCTTAGGCTGGCGACCGGATTGCGGAACATACCTTCTATCCAGAGTCGCGGGC TCTTCGGAGCTATCGCGGGCTTTATAGAGGGAGGATGGACTGGAATGGTAGATGGATGGTATGGTTACCAACAT CAGAACGAACAGGGATCCGGGTACGCAGCAGATTTGAAATCAACACAGAACGCCATCGACGAGATCACCAATAA GGTGAAGTCTGTAATTGAAAAATGAATACGCAATTCAGTGCAGTGGGGAAGGAATTCACCATCTGGAGAAAC GAATTGAAAACCTTAACAAGAAGGTAGATGACGGTTTCTCGATATCTGGACATATAATGCAGAACTTTTGGTA TTGCTGGAAAATGAACGGACCTGGATTATCACGATTCAAACGTTAAAAATCTCTATGAGAAGGTTGATCTCA ACTGAAGAACAACGCCAAGGAAATAGGAAACGGATGTTTCGAGTTCTATCATAAATGCGATAACACATGCATGG AGAGCGTCAAGAACGGTACCTACGACTATCCTAAGTATAGTGAGGAAGCCAACTCAATAGGGAAGAGATCGAC GGAGTCAAATTGGAATCAACGCAATATATCAGATTCTTGCAATTTACAGCACTGTGCGGAGTAGCCTTGTGTT GGTTGTGAGCCTCGGTGCTATTTCTTTTGGATGTGCTCAAACGGCTCTCTCCAGTGTAGAATCTGCATTTGA
Influenza A H7N9 Anhui/1/13 HA

ATGAATACCCAAATATTGGTATTTCGCTCTGATTGCAATTATTCGACTAACGCAGATAAGATTTGCCTTGGCCA
CCATGCTGTGAGTAATGGAACATAAGTTAACACACTTACCGAACGGGGCGTTGAAGTCGTGAACGCCACAGAGA
CAGTCGAGAGAACAAACATTCCACGAATATGCAGCAAAGGCAAACGAACTGTAGACCTCGGGCAATGCGGCCCTC
CTCGGTACCATTACCGGTCCCTCCGCAGTGTGACCAGTTTCTGGAGTTTTTCAGCCGACCTCATCATTGAGCGACG
GGAGGGAAGCGACGTCTGCTACCCTGGGAAGTTTCGTGAACGAAGAAGCGCTGCGGCAGATTTTGAGAGAAAGTG
GTGGTATAGATAAAGAGGCGATGGGCTTCACGTATTCGGGTATAAGGACAAATGGTGCCACGTCTGCATGCAGG
CGGAGCGGCAGCTCTTTTTACGCAGAGATGAAATGGCTGTTGTCAAACACCGATGATGCGGCTTTTCTCAAAT
GACCAAAAGCTATAAGAACACCAGGAAATCCCCGCACTCATAGTCTGGGGTATACATCACTCAGTGTCCACAG
CAGAACAAACGAACTTTATGGGTCTGGCAACAACTGGTGACAGTGGGGTCCTCTAACTATCAACAAAGCTTT
GTGCCATCACCAGGGGCTCGACCACAAGTAAACGGACTCAGTGGGCGGATCGACTTCCACTGGTTGATGCTCAA
TCCTAACGATACCGTTACCTTCTCTTCAACGGAGCCTTCATAGCGCCTGACAGGGCCAGCTTCTCAGGGGTA
AATCCATGGGGATACAGTCAGGGGTCCAGGTGGATGCCAATTGCGAAGGCGATTGCTATCACTCTGGCGGAACA
ATAATCTCCAATCTCCCGTTCCAGAATATTGACTCACGGGCAGTAGGGAAATGTCCCGCTATGTCAAACAGAG
GAGCTTGCTGCTTGCAACCGGCATGAAGAACGTGCCTGAAATACCCAAAGGTAGGGGCCTTTTCGGGGCTATCG
CGGGATTTATCGAAAACGGGTGGGAGGGACTCATCGACGGCTGGTACGGCTTTAGGCATCAAAACGCGCAAGGT
GAAGGCACGGCAGCTGACTACAAGAGCACGCAGTCTGCCATCGATCAGATAACCGGGAACTTAATCGCCTGAT
CGAAAAGACAAATCAGCAATTTGAACTCATCGACAATGAATTTAACGAAGTGGAGAAGCAAATTGGCAATGTCA
TCAACTGGACAAGAGACTCAATTACGGAAGTTTGGAGCTACAATGCTGAATTGCTTGTAGCAATGGAAAACCAG
CATACGATAGACCTGGCTGATTCTGAGATGGACAAGCTCTATGAGCGGGTAAAAAGGCAGCTCCGAGAAAACGC
CGAGGAGGACGGTACGGGATGCTTCGAGATTTTCCATAAGTGTGACGACGACTGTATGGCAAGTATCCGAAATA
ACACTTACGATCATTCAAATACCGGGAGGAAGCTATGCAAAACAGAATACAGATTGACCCGGTTAAATTGAGC
AGCGGCTATAAAGATGTGATCCTGTGGTTTAGCTTCGGAGCTTCCTGTTTCATTCTTCTGGCAATAGTTATGGG
TCTTGTATTTATTTGTGTAAAAACGGGAATATGCGATGCACGATCTGTATCTGA

REFERENCES

1. Desrichard, A., Snyder, A. & Chan, T. A. Cancer Neoantigens and Applications for Immunotherapy. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* (2015). doi:10.1158/1078-0432.CCR-14-3175
2. Schumacher, T. N. & Schreiber, R. D. Neoantigens in cancer immunotherapy. *Science* **348**, 69–74 (2015).
3. Gubin, M. M., Artyomov, M. N., Mardis, E. R. & Schreiber, R. D. Tumor neoantigens: building a framework for personalized cancer immunotherapy. *J. Clin. Invest.* **125**, 3413–3421 (2015).
4. Rizvi, N. A. *et al.* Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science* **348**, 124–128 (2015).
5. Snyder, A. *et al.* Genetic basis for clinical response to CTLA-4 blockade in melanoma. *N. Engl. J. Med.* **371**, 2189–2199 (2014).
6. Carreno, B. M. *et al.* Cancer immunotherapy. A dendritic cell vaccine increases the breadth and diversity of melanoma neoantigen-specific T cells. *Science* **348**, 803–808 (2015).
7. Tran, E. *et al.* Cancer immunotherapy based on mutation-specific CD4⁺ T cells in a patient with epithelial cancer. *Science* **344**, 641–645 (2014).
8. Hacothen, N. & Wu, C. J.-Y. United States Patent Application: 20110293637 - COMPOSITIONS AND METHODS OF IDENTIFYING TUMOR SPECIFIC NEOANTIGENS. (A1). at <<http://appft1.uspto.gov/netacgi/nph-Parser?Sect1=PTO1&Sect2=HITOFF&d=PG01&p=1&u=/netahtml/PTO/srchnum.html&r=1&f=G&l=50&s1=20110293637.PG&NR.>>>
9. Lundegaard, C., Hoof, I., Lund, O. & Nielsen, M. State of the art and challenges in sequence based T-cell epitope prediction. *Immunome Res.* **6 Suppl 2**, S3 (2010).
10. Yadav, M. *et al.* Predicting immunogenic tumour mutations by combining mass spectrometry and exome sequencing. *Nature* **515**, 572–576 (2014).
11. Bassani-Sternberg, M., Pletscher-Frankild, S., Jensen, L. J. & Mann, M. Mass spectrometry of human leukocyte antigen class I peptidomes reveals strong effects of protein abundance and turnover on antigen presentation. *Mol. Cell. Proteomics MCP* **14**, 658–673 (2015).
12. Van Allen, E. M. *et al.* Genomic correlates of response to CTLA-4 blockade in metastatic melanoma. *Science* **350**, 207–211 (2015).
13. Yoshida, K. & Ogawa, S. Splicing factor mutations and cancer. *Wiley Interdiscip. Rev. RNA* **5**, 445–459 (2014).
14. Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma. *Nature* **511**, 543–550 (2014).
15. Rajasagi, M. *et al.* Systematic identification of personal tumor-specific neoantigens in chronic lymphocytic leukemia. *Blood* **124**, 453–462 (2014).
16. Downing, S. R. *et al.* United States Patent Application: 0120208706 - OPTIMIZATION OF MULTIGENE ANALYSIS OF TUMOR SAMPLES. (A1). at <<http://appft1.uspto.gov/netacgi/nph-Parser?Sect1=PTO1&Sect2=HITOFF&d=PG01&p=1&u=/netahtml/PTO/srchnum.html&r=1&f=G&l=50&s1=20120208706.PG&NR.>>>
17. Target Capture for NextGen Sequencing - IDT. at <<http://www.idtdna.com/pages/products/nextgen/target-capture>>
18. Shukla, S. A. *et al.* Comprehensive analysis of cancer-associated somatic mutations in class I HLA genes. *Nat. Biotechnol.* **33**, 1152–1158 (2015).

19. Cieslik, M. *et al.* The use of exome capture RNA-seq for highly degraded RNA with application to clinical cancer sequencing. *Genome Res.* **25**, 1372–1381 (2015).
20. Bodini, M. *et al.* The hidden genomic landscape of acute myeloid leukemia: subclonal structure revealed by undetected mutations. *Blood* **125**, 600–605 (2015).
21. Saunders, C. T. *et al.* Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. *Bioinforma. Oxf. Engl.* **28**, 1811–1817 (2012).
22. Cibulskis, K. *et al.* Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat. Biotechnol.* **31**, 213–219 (2013).
23. Wilkerson, M. D. *et al.* Integrated RNA and DNA sequencing improves mutation detection in low purity tumors. *Nucleic Acids Res.* **42**, e107 (2014).
24. Mose, L. E., Wilkerson, M. D., Hayes, D. N., Perou, C. M. & Parker, J. S. ABRA: improved coding indel detection via assembly-based realignment. *Bioinforma. Oxf. Engl.* **30**, 2813–2815 (2014).
25. Ye, K., Schulz, M. H., Long, Q., Apweiler, R. & Ning, Z. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinforma. Oxf. Engl.* **25**, 2865–2871 (2009).
26. Lam, H. Y. K. *et al.* Nucleotide-resolution analysis of structural variants using BreakSeq and a breakpoint library. *Nat. Biotechnol.* **28**, 47–55 (2010).
27. Frampton, G. M. *et al.* Development and validation of a clinical cancer genomic profiling test based on massively parallel DNA sequencing. *Nat. Biotechnol.* **31**, 1023–1031 (2013).
28. Boegel, S. *et al.* HLA typing from RNA-Seq sequence reads. *Genome Med.* **4**, 102 (2012).
29. Liu, C. *et al.* ATHLATES: accurate typing of human leukocyte antigen through exome sequencing. *Nucleic Acids Res.* **41**, e142 (2013).
30. Mayor, N. P. *et al.* HLA Typing for the Next Generation. *PloS One* **10**, e0127153 (2015).
31. Roy, C. K., Olson, S., Graveley, B. R., Zamore, P. D. & Moore, M. J. Assessing long-distance RNA sequence connectivity via RNA-templated DNA-DNA ligation. *eLife* **4**, (2015).
32. Song, L. & Florea, L. CLASS: constrained transcript assembly of RNA-seq reads. *BMC Bioinformatics* **14 Suppl 5**, S14 (2013).
33. Maretty, L., Sibbesen, J. A. & Krogh, A. Bayesian transcriptome assembly. *Genome Biol.* **15**, 501 (2014).
34. Pertea, M. *et al.* StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **33**, 290–295 (2015).
35. Roberts, A., Pimentel, H., Trapnell, C. & Pachter, L. Identification of novel transcripts in annotated genomes using RNA-Seq. *Bioinforma. Oxf. Engl.* (2011). doi:10.1093/bioinformatics/btr355
36. Vitting-Seerup, K., Porse, B. T., Sandelin, A. & Waage, J. spliceR: an R package for classification of alternative splicing and prediction of coding potential from RNA-seq data. *BMC Bioinformatics* **15**, 81 (2014).
37. Rivas, M. A. *et al.* Human genomics. Effect of predicted protein-truncating genetic variants on the human transcriptome. *Science* **348**, 666–669 (2015).
38. Skelly, D. A., Johansson, M., Madeoy, J., Wakefield, J. & Akey, J. M. A powerful and flexible statistical framework for testing hypotheses of allele-specific gene expression from RNA-seq data. *Genome Res.* **21**, 1728–1737 (2011).
39. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinforma. Oxf. Engl.* **31**, 166–169 (2015).

40. Furney, S. J. *et al.* SF3B1 mutations are associated with alternative splicing in uveal melanoma. *Cancer Discov.* (2013). doi:10.1158/2159-8290.CD-13-0330
41. Zhou, Q. *et al.* A chemical genetics approach for the functional assessment of novel cancer genes. *Cancer Res.* (2015). doi:10.1158/0008-5472.CAN-14-2930
42. Maguire, S. L. *et al.* SF3B1 mutations constitute a novel therapeutic target in breast cancer. *J. Pathol.* **235**, 571–580 (2015).
43. Carithers, L. J. *et al.* A Novel Approach to High-Quality Postmortem Tissue Procurement: The GTEx Project. *Biopreservation Biobanking* **13**, 311–319 (2015).
44. Xu, G. *et al.* RNA CoMPASS: a dual approach for pathogen and host transcriptome analysis of RNA-seq datasets. *PloS One* **9**, e89445 (2014).
45. Andreatta, M. & Nielsen, M. Gapped sequence alignment using artificial neural networks: application to the MHC class I system. *Bioinforma. Oxf. Engl.* (2015). doi:10.1093/bioinformatics/btv639
46. Jørgensen, K. W., Rasmussen, M., Buus, S. & Nielsen, M. NetMHCstab - predicting stability of peptide-MHC-I complexes; impacts for cytotoxic T lymphocyte epitope discovery. *Immunology* **141**, 18–26 (2014).
47. Larsen, M. V. *et al.* An integrative approach to CTL epitope prediction: a combined algorithm integrating MHC class I binding, TAP transport efficiency, and proteasomal cleavage predictions. *Eur. J. Immunol.* **35**, 2295–2303 (2005).
48. Nielsen, M., Lundegaard, C., Lund, O. & Keşmir, C. The role of the proteasome in generating cytotoxic T-cell epitopes: insights obtained from improved predictions of proteasomal cleavage. *Immunogenetics* **57**, 33–41 (2005).
49. Boisvert, F.-M. *et al.* A Quantitative Spatial Proteomics Analysis of Proteome Turnover in Human Cells. *Mol. Cell. Proteomics* **11**, M111.011429–M111.011429 (2012).
50. Duan, F. *et al.* Genomic and bioinformatic profiling of mutational neoepitopes reveals new rules to predict anticancer immunogenicity. *J. Exp. Med.* **211**, 2231–2248 (2014).
51. Janeway's Immunobiology: 9780815345312: Medicine & Health Science Books @ Amazon.com. at <<http://www.amazon.com/Janeways-Immunobiology-Kenneth-Murphy/dp/0815345313>>
52. Calis, J. J. A. *et al.* Properties of MHC Class I Presented Peptides That Enhance Immunogenicity. *PLoS Comput. Biol.* **9**, e1003266 (2013).
53. Zhang, J. *et al.* Intratumor heterogeneity in localized lung adenocarcinomas delineated by multiregion sequencing. *Science* **346**, 256–259 (2014)
54. Walter, M. J. *et al.* Clonal architecture of secondary acute myeloid leukemia. *N. Engl. J. Med.* 366, 1090–1098 (2012).
55. Hunt DF, Henderson RA, Shabanowitz J, Sakaguchi K, Michel H, Sevilir N, Cox AL, Appella E, Engelhard VH. Characterization of peptides bound to the class I MHC molecule HLA-A2.1 by mass spectrometry. *Science* 1992. 255: 1261-1263.
56. Zarling AL, Polefrone JM, Evans AM, Mikesch LM, Shabanowitz J, Lewis ST, Engelhard VH, Hunt DF. Identification of class I MHC-associated phosphopeptides as targets for cancer immunotherapy. *Proc Natl Acad Sci U S A.* 2006 Oct 3;103(40):14889-94.
57. Bassani-Sternberg M, Pletscher-Frankild S, Jensen LJ, Mann M. Mass spectrometry of human leukocyte antigen class I peptidomes reveals strong effects of protein abundance and turnover on antigen presentation. *Mol Cell Proteomics.* 2015 Mar;14(3):658-73. doi: 10.1074/mcp.M114.042812.
58. Abelin JG, Trantham PD, Penny SA, Patterson AM, Ward ST, Hildebrand WH, Cobbold M, Bai DL, Shabanowitz J, Hunt DF. Complementary IMAC enrichment methods for HLA-associated phosphopeptide identification by mass spectrometry. *Nat Protoc.* 2015 Sep;10(9):1308-18. doi: 10.1038/nprot.2015.086. Epub 2015 Aug 6

59. Barnstable CJ, Bodmer WF, Brown G, Galfre G, Milstein C, Williams AF, Ziegler A. Production of monoclonal antibodies to group A erythrocytes, HLA and other human cell surface antigens-new tools for genetic analysis. *Cell*. 1978 May;14(1):9-20.
60. Goldman JM, Hibbin J, Kearney L, Orchard K, Th'ng KH. HLA-DR monoclonal antibodies inhibit the proliferation of normal and chronic granulocytic leukaemia myeloid progenitor cells. *Br J Haematol*. 1982 Nov;52(3):411-20.
61. Eng JK, Jahan TA, Hoopmann MR. Comet: an open-source MS/MS sequence database search tool. *Proteomics*. 2013 Jan;13(1):22-4. doi: 10.1002/pmic.201200439. Epub 2012 Dec 4.
62. Eng JK, Hoopmann MR, Jahan TA, Egertson JD, Noble WS, MacCoss MJ. A deeper look into Comet--implementation and features. *J Am Soc Mass Spectrom*. 2015 Nov;26(11):1865-74. doi: 10.1007/s13361-015-1179-x. Epub 2015 Jun 27.
63. Lukas Käll, Jesse Canterbury, Jason Weston, William Stafford Noble and Michael J. MacCoss. Semi-supervised learning for peptide identification from shotgun proteomics datasets. *Nature Methods* 4:923 – 925, November 2007
64. Lukas Käll, John D. Storey, Michael J. MacCoss and William Stafford Noble. Assigning confidence measures to peptides identified by tandem mass spectrometry. *Journal of Proteome Research*, 7(1):29-34, January 2008
65. Lukas Käll, John D. Storey and William Stafford Noble. Nonparametric estimation of posterior error probabilities associated with peptides identified by tandem mass spectrometry. *Bioinformatics*, 24(16):i42-i48, August 2008
66. Kinney RM, BJ Johnson , VL Brown , DW Trent. Nucleotide Sequence of the 26 S mRNA of the Virulent Trinidad Donkey Strain of Venezuelan Equine Encephalitis Virus and Deduced Sequence of the Encoded Structural Proteins. *Virology* 152 (2), 400-413. 1986 Jul 30.
67. Jill E Slansky, Frédérique M Rattis, Lisa F Boyd, Tarek Fahmy, Elizabeth M Jaffee, Jonathan P Schneck, David H Margulies, Drew M Pardoll. Enhanced Antigen-Specific Antitumor Immunity with Altered Peptide Ligands that Stabilize the MHC-Peptide-TCR Complex. *Immunity*, Volume 13, Issue 4, 1 October 2000, Pages 529-538.
68. A Y Huang, P H Gulden, A S Woods, M C Thomas, C D Tong, W Wang, V H Engelhard, G Pasternack, R Cotter, D Hunt, D M Pardoll, and E M Jaffee. The immunodominant major histocompatibility complex class I-restricted antigen of a murine colon tumor derives from an endogenous retroviral gene product. *Proc Natl Acad Sci U S A.*; 93(18): 9730–9735, 1996 Sep 3.
69. JOHNSON, BARBARA J. B., RICHARD M. KINNEY, CRYSTLE L. KOST AND DENNIS W. TRENT. Molecular Determinants of Alphavirus Neurovirulence: Nucleotide and Deduced Protein Sequence Changes during Attenuation of Venezuelan Equine Encephalitis Virus. *J Gen Virol* 67:1951-1960, 1986.
70. Aarnoudse, C.A., Krüse, M., Konopitzky, R., Brouwenstijn, N., and Schrier, P.I. (2002). TCR reconstitution in Jurkat reporter cells facilitates the identification of novel tumor antigens by cDNA expression cloning. *Int J Cancer* 99, 7–13.
71. Alexander, J., Sidney, J., Southwood, S., Ruppert, J., Oseroff, C., Maewal, A., Snoke, K., Serra, H.M., Kubo, R.T., and Sette, A. (1994). Development of high potency universal DR-restricted helper epitopes by modification of high affinity DR-blocking peptides. *Immunity* 1, 751–761.
72. Banu, N., Chia, A., Ho, Z.Z., Garcia, A.T., Paravasivam, K., Grotenbreg, G.M., Bertoletti, A., and Gehring, A.J. (2014). Building and optimizing a virus-specific T cell receptor library for targeted immunotherapy in viral infections. *Scientific Reports* 4, 4166.
73. Cornet, S., Miconnet, I., Menez, J., Lemonnier, F., and Kosmatopoulos, K. (2006). Optimal organization of a polypeptide-based candidate cancer vaccine composed of cryptic

tumor peptides with enhanced immunogenicity. *Vaccine* 24, 2102–2109.

74. Depla, E., van der Aa, A., Livingston, B.D., Crimi, C., Allosery, K., de Brabandere, V., Krakover, J., Murthy, S., Huang, M., Power, S., et al. (2008). Rational design of a multiepitope vaccine encoding T-lymphocyte epitopes for treatment of chronic hepatitis B virus infections. *Journal of Virology* 82, 435–450.

75. Ishioka, G.Y., Fikes, J., Hermanson, G., Livingston, B., Crimi, C., Qin, M., del Guercio, M.F., Oseroff, C., Dahlberg, C., Alexander, J., et al. (1999). Utilization of MHC class I transgenic mice for development of minigene DNA vaccines encoding multiple HLA-restricted CTL epitopes. *J Immunol* 162, 3915–3925.

76. Janetzki, S., Price, L., Schroeder, H., Britten, C.M., Welters, M.J.P., and Hoos, A. (2015). Guidelines for the automated evaluation of Elispot assays. *Nat Protoc* 10, 1098–1115.

77. Lyons, G.E., Moore, T., Brasic, N., Li, M., Roszkowski, J.J., and Nishimura, M.I. (2006). Influence of human CD8 on antigen recognition by T-cell receptor-transduced cells. *Cancer Res* 66, 11455–11461.

78. Nagai, K., Ochi, T., Fujiwara, H., An, J., Shirakata, T., Mineno, J., Kuzushima, K., Shiku, H., Melenhorst, J.J., Gostick, E., et al. (2012). Aurora kinase A-specific T-cell receptor gene transfer redirects T lymphocytes to display effective antileukemia reactivity. *Blood* 119, 368–376.

79. Panina-Bordignon, P., Tan, A., Termijtelen, A., Demotz, S., Corradin, G., and Lanzavecchia, A. (1989). Universally immunogenic T cell epitopes: promiscuous binding to human MHC class II and promiscuous recognition by T cells. *Eur J Immunol* 19, 2237–2242.

80. Vitiello, A., Marchesini, D., Furze, J., Sherman, L.A., and Chesnut, R.W. (1991). Analysis of the HLA-restricted influenza-specific cytotoxic T lymphocyte response in transgenic mice carrying a chimeric human-mouse class I major histocompatibility complex. *J Exp Med* 173, 1007–1015.

81. Yachi, P.P., Ampudia, J., Zal, T., and Gascoigne, N.R.J. (2006). Altered peptide ligands induce delayed CD8-T cell receptor interaction—a role for CD8 in distinguishing antigen quality. *Immunity* 25, 203–211.

82. Pushko P, Parker M, Ludwig GV, Davis NL, Johnston RE, Smith JF. Replicon-helper systems from attenuated Venezuelan equine encephalitis virus: expression of heterologous genes in vitro and immunization against heterologous pathogens in vivo. *Virology*. 1997 Dec 22;239(2):389-401.

83. Strauss, JH and E G Strauss. The alphaviruses: gene expression, replication, and evolution. *Microbiol Rev*. 1994 Sep; 58(3): 491–562.

84. Rhône C, Ehrengreuer MU, Grandgirard D. Alphaviral cytotoxicity and its implication in vector development. *Exp Physiol*. 2005 Jan;90(1):45-52. Epub 2004 Nov 12.

85. Riley, Michael K. II, and Wilfred Vermerris. Recent Advances in Nanomaterials for Gene Delivery—A Review. *Nanomaterials* 2017, 7(5), 94.

86. Frolov I, Hardy R, Rice CM. Cis-acting RNA elements at the 5' end of Sindbis virus genome RNA regulate minus- and plus-strand RNA synthesis. *RNA*. 2001 Nov;7(11):1638-51.

87. Jose J, Snyder JE, Kuhn RJ. A structural and functional perspective of alphavirus replication and assembly. *Future Microbiol*. 2009 Sep;4(7):837-56.

88. Bo Li and C. olin N. Dewey. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*, 12:323, August 2011

89. Hillary Pearson, Tariq Daouda, Diana Paola Granados, Chantal Durette, Eric Bonneil, Mathieu Courcelles, Anja Rodenbrock, Jean-Philippe Laverdure, Caroline Côté, Sylvie Mader, Sébastien Lemieux, Pierre Thibault, and Claude Perreault. MHC class I-associated peptides derive from selective regions of the human genome. *The Journal of Clinical Investigation*, 2016,

90. Juliane Liepe, Fabio Marino, John Sidney, Anita Jeko, Daniel E. Bunting, Alessandro Sette, Peter M. Klotzel, Michael P. H. Stumpf, Albert J. R. Heck, Michele Mishto. A large fraction of HLA class I ligands are proteasome-generated spliced peptides. *Science*, 21, October 2016.
91. Mommen GP., Marino, F., Meiring HD., Poelen, MC., van Gaans-van den Brink, JA., Mohammed S., Heck AJ., and van Els CA. Sampling From the Proteome to the Human Leukocyte Antigen-DR (HLA-DR) Ligandome Proceeds Via High Specificity. *Mol Cell Proteomics* 15(4): 1412-1423, April 2016.
92. Sebastian Kreiter, Mathias Vormehr, Niels van de Roemer, Mustafa Diken, Martin Löwer, Jan Diekmann, Sebastian Boegel, Barbara Schrörs, Fulvia Vascotto, John C. Castle, Arbel D. Tadmor, Stephen P. Schoenberger, Christoph Huber, Özlem Türeci, and Ugur Sahin. Mutant MHC class II epitopes drive therapeutic immune responses to cancer. *Nature* 520, 692-696, April 2015.
93. Tran E., Turcotte S., Gros A., Robbins P.F., Lu Y.C., Dudley M.E., Wunderlich J.R., Somerville R.P., Hogan K., Hinrichs C.S., Parkhurst M.R., Yang J.C., Rosenberg S.A. Cancer immunotherapy based on mutation-specific CD4⁺ T cells in a patient with epithelial cancer. *Science* 344(6184) 641-645, May 2014.
94. Andreatta M., Karosiene E., Rasmussen M., Stryhn A., Buus S., Nielsen M. Accurate pan-specific prediction of peptide-MHC class II binding affinity with improved binding core identification. *Immunogenetics* 67(11-12) 641-650, November 2015.
95. Nielsen, M., Lund, O. NN-align. An artificial neural network-based alignment algorithm for MHC class II peptide binding prediction. *BMC Bioinformatics* 10:296, September 2009.
96. Nielsen, M., Lundegaard, C., Lund, O. Prediction of MHC class II binding affinity using SMM-align, a novel stabilization matrix alignment method. *BMC Bioinformatics* 8:238, July 2007.
97. Zhang, J., et al. PEAKS DB: de novo sequencing assisted database search for sensitive and accurate peptide identification. *Molecular & Cellular Proteomics*. 11(4):1-8. 1/2/2012.
98. Jensen, Kamilla Kjaergaard, et al. "Improved Methods for Predicting Peptide Binding Affinity to MHC Class II Molecules." *Immunology*, 2018, doi:10.1111/imm.12889.
99. Carter, S.L., Cibulskis, K., Helman, E., McKenna, A., Shen, H., Zack, T., Laird, P.W., Onofrio, R.C., Winckler, W., Weir, B.A., et al. (2012). Absolute quantification of somatic DNA alterations in human cancer. *Nat. Biotechnol.* 30, 413–421
100. McGranahan, N., Rosenthal, R., Hiley, C.T., Rowan, A.J., Watkins, T.B.K., Wilson, G.A., Birkbak, N.J., Veeriah, S., Van Loo, P., Herrero, J., et al. (2017). Allele-Specific HLA Loss and Immune Escape in Lung Cancer Evolution. *Cell* 171, 1259–1271.e11.
101. Shukla, S.A., Rooney, M.S., Rajasagi, M., Tiao, G., Dixon, P.M., Lawrence, M.S., Stevens, J., Lane, W.J., Dellagatta, J.L., Steelman, S., et al. (2015). Comprehensive analysis of cancer-associated somatic mutations in class I HLA genes. *Nat. Biotechnol.* 33, 1152–1158.
102. Van Loo, P., Nordgard, S.H., Lingjærde, O.C., Russnes, H.G., Rye, I.H., Sun, W., Weigman, V.J., Marynen, P., Zetterberg, A., Naume, B., et al. (2010). Allele-specific copy number analysis of tumors. *Proc. Natl. Acad. Sci. U. S. A.* 107, 16910–16915.
103. Van Loo, P., Nordgard, S.H., Lingjærde, O.C., Russnes, H.G., Rye, I.H., Sun, W., Weigman, V.J., Marynen, P., Zetterberg, A., Naume, B., et al. (2010). Allele-specific copy number analysis of tumors. *Proc. Natl. Acad. Sci. U. S. A.* 107, 16910–16915.

CLAIMS

What is claimed is:

1. A composition for delivery of a self-replicating alphavirus-based expression system,

wherein the composition for delivery of the self-replicating alphavirus-based expression system comprises:

(A) the self-replicating alphavirus-based expression system, wherein the self-replicating alphavirus-based expression system comprises one or more vectors, wherein the one or more vectors comprises:

(a) an RNA alphavirus backbone, wherein the RNA alphavirus backbone comprises:

(i) at least one promoter nucleotide sequence, and

(ii) at least one polyadenylation (poly(A)) sequence; and

(b) a cassette, wherein the cassette comprises:

(i) at least one antigen-encoding nucleic acid sequence comprising:

a. a nucleic acid sequence encoding an infectious disease organism peptide selected from the group consisting of: a pathogen-derived peptide, a virus-derived peptide, a bacteria-derived peptide, a fungus-derived peptide, and a parasite-derived peptide,

b. optionally a 5' linker sequence, and

c. optionally a 3' linker sequence;

(ii) optionally, a second promoter nucleotide sequence operably linked to the at least one antigen-encoding nucleic acid sequence; and

(iii) optionally, at least one second poly(A) sequence, wherein the second poly(A) sequence is a native poly(A) sequence or an exogenous poly(A) sequence to the alphavirus, and

(B) a lipid-nanoparticle (LNP), wherein the LNP encapsulates the self-replicating alphavirus-based expression system.

2. A composition for delivery of a self-replicating alphavirus-based expression system,

wherein the composition for delivery of the self-replicating alphavirus-based expression system comprises:

(A) the self-replicating alphavirus-based expression system, wherein the self-replicating alphavirus-based expression system comprises one or more vectors, wherein the one or more vectors comprises:

(a) an RNA alphavirus backbone, wherein the RNA alphavirus backbone comprises the nucleic acid sequence set forth in SEQ ID NO:6, wherein the RNA alphavirus backbone sequence comprises a 26S promoter nucleotide sequence and a poly(A) sequence, wherein the 26S promoter sequence is endogenous to the RNA alphavirus backbone, and wherein the poly(A) sequence is endogenous to the RNA alphavirus backbone; and

(b) a cassette integrated between the 26S promoter nucleotide sequence and the poly(A) sequence, wherein the cassette is operably linked to the 26S promoter nucleotide sequence, and wherein the cassette comprises at least one antigen-encoding nucleic acid sequence comprising:

a. a nucleic acid sequence encoding an infectious disease organism peptides selected from the group consisting of: a pathogen-derived peptide, a virus-derived peptide, a bacteria-derived peptide, a fungus-derived peptide, and a parasite-derived peptide,

b. optionally a 5' linker sequence, and

c. optionally a 3' linker sequence; and

(B) a lipid-nanoparticle (LNP), wherein the LNP encapsulates the self-replicating alphavirus-based expression system.

3. The composition of claims 1 or 2, wherein the nucleic acid sequence encoding the peptide comprises an epitope-encoding nucleic acid sequence.

4. The composition of claims 1 or 2, wherein the nucleic acid sequence encoding the peptide comprises two or more distinct epitope-encoding nucleic acid sequences.
5. The composition of claims 1 or 2, wherein the nucleic acid sequence encoding the peptide comprises two or more distinct epitope-encoding nucleic acid sequences.
6. The composition of claims 1 or 2, wherein the nucleic acid sequence encoding the peptide comprises between 1-10, between 1-20, between 1-30, between 1-40, between 1-50, between 1-100, between 1-200, between 1-300, between 1-400, or between 1-500 distinct epitope-encoding nucleic acid sequences distinct epitope-encoding nucleic acid sequences.
7. The composition of claims 1 or 2, wherein the nucleic acid sequence encoding the peptide comprises between 2-10, between 2-20, between 2-30, between 2-40, between 2-50, between 2-100, between 2-200, between 2-300, between 2-400, or between 2-500 distinct epitope-encoding nucleic acid sequences distinct epitope-encoding nucleic acid sequences.
8. The composition of claims 1 or 2, wherein the nucleic acid sequence encoding the peptide comprises a nucleic acid sequence encoding a full-length protein expressed in the infectious disease organism.
9. The composition of claims 1 or 2, wherein the nucleic acid sequence encoding the peptide comprises a nucleic acid sequence encoding a protein domain of the protein expressed in the infectious disease organism.
10. The composition of claims 1 or 2, wherein the nucleic acid sequence encoding the peptide comprises a nucleic acid sequence encoding a protein subunit of the protein expressed in the infectious disease organism.
11. The composition of claims 1 or 2, wherein the nucleic acid sequence encoding the peptide comprises two or more distinct nucleic acid sequences encoding a peptide selected from: an epitope, a full-length protein, a protein subunit, a protein domain, and combinations thereof of the protein expressed in the infectious disease organism.

12. The composition of any one of claims 1-11, wherein the encoded peptide or peptides is capable of stimulating an immune response when expressed in a subject.
13. The composition of any one of claims 1-11, wherein the encoded peptide or peptides is capable of stimulating a T cell response when expressed in a subject .
14. The composition of any one of claims 1-11, wherein the encoded peptide or peptides is capable of stimulating a B cell response when expressed in a subject .
15. The composition of any one of claims 1-11, wherein the encoded peptide or peptides is capable of stimulating a T cell response and a B cell response when expressed in a subject.
16. The composition of any one of claims 1-15, wherein the infectious disease organism is selected from the group consisting of: Severe acute respiratory syndrome-related coronavirus (SARS), severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), Ebola, HIV, Hepatitis B virus (HBV), influenza, Hepatitis C virus (HCV), Human papillomavirus (HPV), Cytomegalovirus (CMV), Chikungunya virus, Respiratory syncytial virus (RSV), Dengue virus, a orthomyxoviridae family virus, and tuberculosis.
17. The composition of any one of claims 1, or 3-16, wherein an ordered sequence of each element of the cassette in the composition for delivery of the self-replicating alphavirus-based expression system is described in the formula, from 5' to 3', comprising

$$P_a-(L5_b-N_c-L3_d)_x-(G5_e-U_f)_y-G3_g$$

wherein P comprises the second promoter nucleotide sequence, where a = 0 or 1,

N comprises one of the epitope-encoding nucleic acid sequences, wherein the epitope-encoding nucleic acid sequence comprises an MHC class I epitope-encoding nucleic acid sequence, where c = 1,

L5 comprises the 5' linker sequence, where b = 0 or 1,

L3 comprises the 3' linker sequence, where d = 0 or 1,

G5 comprises one of the at least one nucleic acid sequences encoding a GPGPG amino acid linker, where $e = 0$ or 1 ,

G3 comprises one of the at least one nucleic acid sequences encoding a GPGPG amino acid linker, where $g = 0$ or 1 ,

U comprises one of the at least one MHC class II epitope-encoding nucleic acid sequence, where $f = 1$,

$X = 1$ to 400 , where for each X the corresponding N_c is an MHC class I epitope-encoding nucleic acid sequence, and

$Y = 0, 1$, or 2 , where for each Y the corresponding U_f is an MHC class II epitope-encoding nucleic acid sequence.

18. The composition of claim 17, wherein for each X the corresponding N_c is a distinct MHC class I epitope-encoding nucleic acid sequence.
19. The composition of claim 17 or 18, wherein for each Y the corresponding U_f is a distinct MHC class II epitope-encoding nucleic acid sequence.
20. The composition of any one of claims 17-19, wherein

$$a = 0, b = 1, d = 1, e = 1, g = 1, h = 1, X = 20, Y = 2,$$

the at least one promoter nucleotide sequence is a single 26S promoter nucleotide sequence provided by the RNA alphavirus backbone,

the at least one polyadenylation poly(A) sequence is a poly(A) sequence of at least 100 consecutive A nucleotides provided by the RNA alphavirus backbone,

the cassette is integrated between the 26S promoter nucleotide sequence and the poly(A) sequence, wherein the cassette is operably linked to the 26S promoter nucleotide sequence and the poly(A) sequence,

each N encodes a MHC class I epitope 7-15 amino acids in length,

L5 is a native 5' linker sequence that encodes a native N-terminal amino acid sequence of the MHC I epitope, and wherein the 5' linker sequence encodes a peptide that is at least 3 amino acids in length,

L3 is a native 3' linker sequence that encodes a native C-terminal amino acid sequence of the MHC I epitope, and wherein the 3' linker sequence encodes a peptide that is at least 3 amino acids in length,

U is each of a PADRE class II sequence and a Tetanus toxoid MHC class II sequence,

the RNA alphavirus backbone is the sequence set forth in SEQ ID NO:6, and

each of the MHC class I epitope-encoding nucleic acid sequences encodes a polypeptide that is between 13 and 25 amino acids in length.

21. The composition of any of the above claims, wherein the LNP comprises a lipid selected from the group consisting of: an ionizable amino lipid, a phosphatidylcholine, cholesterol, a PEG-based coat lipid, or a combination thereof.
22. The composition of any of the above claims, wherein the LNP comprises an ionizable amino lipid, a phosphatidylcholine, cholesterol, and a PEG-based coat lipid.
23. The composition of claim 21 or 22, wherein the ionizable amino lipids comprise MC3-like (dilinoleylmethyl-4-dimethylaminobutyrate) molecules.
24. The composition of any of the above claims, wherein the LNP-encapsulated expression system has a diameter of about 100nm.
25. The composition of any of the above claims, wherein the composition for delivery of the self-replicating alphavirus-based expression system is formulated for intramuscular (IM), intradermal (ID), subcutaneous (SC), or intravenous (IV) administration.
26. The composition of any of the above claims, wherein the composition for delivery of the self-replicating alphavirus-based expression system is formulated for intramuscular (IM) administration.

27. The composition of any one of claims 1, 3-19, or 21-26, wherein the cassette is integrated between the at least one promoter nucleotide sequence and the at least one poly(A) sequence.
28. The composition of any one of claims 1, 3-19, or 21-27, wherein the at least one promoter nucleotide sequence is operably linked to the cassette.
29. The composition of any one of claims 1, 3-19, or 21-28, wherein the one or more vectors comprise one or more +-stranded RNA vectors.
30. The composition of claim 29 wherein the one or more +-stranded RNA vectors comprise a 5' 7-methylguanosine (m7g) cap.
31. The composition of claim 29 or 30, wherein the one or more +-stranded RNA vectors are produced by *in vitro* transcription.
32. The composition of any one of claims 1, 3-19, or 21-31, wherein the one or more vectors are self-replicating within a mammalian cell.
33. The composition of any one of 1, 3-19, or 21-32, wherein the RNA alphavirus backbone comprises at least one nucleotide sequence of an Aura virus, a Fort Morgan virus, a Venezuelan equine encephalitis virus, a Ross River virus, a Semliki Forest virus, a Sindbis virus, or a Mayaro virus.
34. The composition of any one of claims 1, 3-19, or 21-32, wherein the RNA alphavirus backbone comprises at least one nucleotide sequence of a Venezuelan equine encephalitis virus.
35. The composition of claim 33 or 34, wherein the RNA alphavirus backbone comprises at least sequences for nonstructural protein-mediated amplification, a 26S promoter sequence, a poly(A) sequence, a nonstructural protein 1 (nsP1) gene, a nsP2 gene, a nsP3 gene, and a nsP4 gene encoded by the nucleotide sequence of the Aura virus, the Fort Morgan virus, the Venezuelan equine encephalitis virus, the Ross River virus, the Semliki Forest virus, the Sindbis virus, or the Mayaro virus.
36. The composition of claim 33 or 34, wherein the RNA alphavirus backbone comprises at least sequences for nonstructural protein-mediated amplification, a 26S promoter

sequence, and a poly(A) sequence encoded by the nucleotide sequence of the Aura virus, the Fort Morgan virus, the Venezuelan equine encephalitis virus, the Ross River virus, the Semliki Forest virus, the Sindbis virus, or the Mayaro virus.

37. The composition of claim 35 or 36, wherein sequences for nonstructural protein-mediated amplification are selected from the group consisting of: an alphavirus 5' UTR, a 51-nt CSE, a 24-nt CSE, a 26S subgenomic promoter sequence, a 19-nt CSE, an alphavirus 3' UTR, or combinations thereof.
38. The composition of any one of claims 35-37, wherein the RNA alphavirus backbone does not encode structural virion proteins capsid, E2 and E1.
39. The composition of claim 38, wherein the cassette is inserted in place of structural virion proteins within the nucleotide sequence of the Aura virus, the Fort Morgan virus, the Venezuelan equine encephalitis virus, the Ross River virus, the Semliki Forest virus, the Sindbis virus, or the Mayaro virus.
40. The composition of claim 33 or 34, wherein the Venezuelan equine encephalitis virus comprises the sequence of SEQ ID NO:3 or SEQ ID NO:5.
41. The composition of claim 33 or 34, wherein the Venezuelan equine encephalitis virus comprises the sequence of SEQ ID NO:3 or SEQ ID NO:5 further comprising a deletion between base pair 7544 and 11175.
42. The composition of claim 41, wherein the RNA alphavirus backbone comprises the sequence set forth in SEQ ID NO:6 or SEQ ID NO:7.
43. The composition of claim 41 or 42, wherein the cassette is inserted at position 7544 to replace the deletion between base pairs 7544 and 11175 as set forth in the sequence of SEQ ID NO:3 or SEQ ID NO:5.
44. The composition of claim 39-43, wherein the insertion of the cassette provides for transcription of a polycistronic RNA comprising the nsP1-4 genes and the at least one nucleic acid sequence, wherein the nsP1-4 genes and the at least one nucleic acid sequence are in separate open reading frames.

45. The composition of any one of claims 1, 3-19, or 21-44, wherein the at least one promoter nucleotide sequence is the native 26S promoter nucleotide sequence encoded by the RNA alphavirus backbone.
46. The composition of any one of claims 1, 3-19, or 21-44, wherein the at least one promoter nucleotide sequence is an exogenous RNA promoter.
47. The composition of any one of claims 1, 3-19, or 21-46, wherein the second promoter nucleotide sequence is a 26S promoter nucleotide sequence.
48. The composition of any one of claims 1, 3-19, or 21-46, wherein the second promoter nucleotide sequence comprises multiple 26S promoter nucleotide sequences, wherein each 26S promoter nucleotide sequence provides for transcription of one or more of the separate open reading frames.
49. The composition of any one of the above claims, wherein the one or more vectors are each at least 300nt in size.
50. The composition of any one of the above claims, wherein the one or more vectors are each at least 1kb in size.
51. The composition of any one of the above claims, wherein the one or more vectors are each 2kb in size.
52. The composition of any one of the above claims, wherein the one or more vectors are each less than 5kb in size.
53. The composition of any one of claims 1-19, or 21-52, wherein the at least one antigen-encoding nucleic acid sequence comprises two or more antigen-encoding nucleic acid sequences.
54. The composition of claim 53, wherein each antigen-encoding nucleic acid sequence is linked directly to one another.
55. The composition of any one of claims 1-19, or 21-54, wherein each antigen-encoding nucleic acid sequence is linked to a distinct antigen-encoding nucleic acid sequence with a nucleic acid sequence encoding a linker.

56. The composition of claim 55, wherein the linker links two MHC class I epitope-encoding nucleic acid sequences or an MHC class I epitope-encoding nucleic acid sequence to an MHC class II epitope-encoding nucleic acid sequence.
57. The composition of claim 56, wherein the linker is selected from the group consisting of: (1) consecutive glycine residues, at least 2, 3, 4, 5, 6, 7, 8, 9, or 10 residues in length; (2) consecutive alanine residues, at least 2, 3, 4, 5, 6, 7, 8, 9, or 10 residues in length; (3) two arginine residues (RR); (4) alanine, alanine, tyrosine (AAY); (5) a consensus sequence at least 2, 3, 4, 5, 6, 7, 8, 9, or 10 amino acid residues in length that is processed efficiently by a mammalian proteasome; and (6) one or more native sequences flanking the antigen derived from the cognate protein of origin and that is at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, or 2-20 amino acid residues in length.
58. The composition of claim 55, wherein the linker links two MHC class II epitope-encoding nucleic acid sequences or an MHC class II sequence to an MHC class I epitope-encoding nucleic acid sequence.
59. The composition of claim 58, wherein the linker comprises the sequence GPGPG.
60. The composition of any one of claims 1-19, or 21-59, wherein the antigen-encoding nucleic acid sequences is linked, operably or directly, to a separate or contiguous sequence that enhances the expression, stability, cell trafficking, processing and presentation, and/or immunogenicity of the antigen-encoding nucleic acid sequence.
61. The composition of claim 60, wherein the separate or contiguous sequence comprises at least one of: a ubiquitin sequence, a ubiquitin sequence modified to increase proteasome targeting (e.g., the ubiquitin sequence contains a Gly to Ala substitution at position 76), an immunoglobulin signal sequence (e.g., IgK), a major histocompatibility class I sequence, lysosomal-associated membrane protein (LAMP)-1, human dendritic cell lysosomal-associated membrane protein, and a major histocompatibility class II sequence; optionally wherein the ubiquitin sequence modified to increase proteasome targeting is A76.

62. The composition of any one of claims 1-19, or 21-61, wherein the at least one antigen-encoding nucleic acid sequence comprises at least 2-10, 2, 3, 4, 5, 6, 7, 8, 9, or 10 antigen-encoding nucleic acid sequences, optionally wherein each antigen-encoding nucleic acid sequence encodes a distinct antigen-encoding nucleic acid sequence.
63. The composition of any one of claims 1-19, or 21-61, wherein the at least one antigen-encoding nucleic acid sequence comprises at least 11-20, 15-20, 11-100, 11-200, 11-300, 11-400, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 or up to 400 antigen-encoding nucleic acid sequences, optionally wherein each antigen-encoding nucleic acid sequence encodes a distinct antigen-encoding nucleic acid sequence.
64. The composition any one of claims 1-19, or 21-61, wherein the at least one antigen-encoding nucleic acid sequence comprises at least 11-20, 15-20, 11-100, 11-200, 11-300, 11-400, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 or up to 400 antigen-encoding nucleic acid sequences.
65. The composition of any one of claims 1-19, or 21-61, wherein the at least one antigen-encoding nucleic acid sequence comprises at least 2-400 antigen-encoding nucleic acid sequences and wherein at least two of the antigen-encoding nucleic acid sequences encode epitope sequences or portions thereof that are presented by MHC class I on a cell surface.
66. The composition of 20, wherein at least two of the MHC class I epitopes are presented by MHC class I on the infected cell surface.
67. The composition of any one of claims 1-19, or 21-66, wherein the epitope-encoding nucleic acid sequences comprises at least one MHC class I epitope-encoding nucleic acid sequence, and wherein each antigen-encoding nucleic acid sequence encodes a polypeptide sequence between 8 and 35 amino acids in length, optionally 9-17, 9-25, 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 or 35 amino acids in length.
68. The composition of any one of claims 17-19, or 21-67, wherein the at least one MHC class II epitope-encoding nucleic acid sequence is present.

69. The composition of any one of claims 17-19, or 21-67, wherein the at least one MHC class II epitope-encoding nucleic acid sequence is present and comprises at least one MHC class II epitope-encoding nucleic acid sequence that comprises at least one alteration that makes the encoded epitope sequence distinct from the corresponding peptide sequence encoded by a wild-type nucleic acid sequence.
70. The composition of any one of claims 1-19, or 21-69, wherein the epitope-encoding nucleic acid sequence comprises an MHC class II epitope-encoding nucleic acid sequence and wherein each antigen-encoding nucleic acid sequence encodes a polypeptide sequence that is 12-20, 12, 13, 14, 15, 16, 17, 18, 19, 20, or 20-40 amino acids in length.
71. The composition of any one of claims 1-19, or 21-70, wherein the epitope-encoding nucleic acid sequences comprises an MHC class II epitope-encoding nucleic acid sequence, wherein the at least one MHC class II epitope-encoding nucleic acid sequence is present, and wherein the at least one MHC class II epitope-encoding nucleic acid sequence comprises at least one universal MHC class II epitope-encoding nucleic acid sequence, optionally wherein the at least one universal sequence comprises at least one of Tetanus toxoid and PADRE.
72. The composition of any one of claims 1, 3-19, or 21-71, wherein the at least one promoter nucleotide sequence or the second promoter nucleotide sequence is inducible.
73. The composition of any one of claims 1, 3-19, or 21-71, wherein the at least one promoter nucleotide sequence or the second promoter nucleotide sequence is non-inducible.
74. The composition of any one of claims 1, 3-19, or 21-73, wherein the at least one poly(A) sequence comprises a poly(A) sequence native to the alphavirus.
75. The composition of any one of claims 1, 3-19, or 21-73, wherein the at least one poly(A) sequence comprises a poly(A) sequence exogenous to the alphavirus.
76. The composition of any one of claims 1, 3-19, or 21-75, wherein the at least one poly(A) sequence is operably linked to at least one of the at least one nucleic acid sequences.

77. The composition of any one claims 1, 3-19, or 21-76, wherein the at least one poly(A) sequence is at least 20 , at least 30, at least 40, at least 50, at least 60, at least 70, at least 80, or at least 90 consecutive A nucleotides.
78. The composition of any one of claims 1, 3-19, or 21-76, wherein the at least one poly(A) sequence is at least 100 consecutive A nucleotides.
79. The composition of any one of claims 1-19, or 21-78, wherein the epitope-encoding nucleic acid sequence comprises a MHC class I epitope-encoding nucleic acid sequence, and wherein the MHC class I epitope-encoding nucleic acid sequence is selected by performing the steps of:
- (a) obtaining at least one of exome, transcriptome, or whole genome nucleotide sequencing data from the infectious disease organism, wherein the infectious disease organism nucleotide sequencing data is used to obtain data representing peptide sequences of each of a set of epitopes;
 - (b) inputting the peptide sequence of each epitope into a presentation model to generate a set of numerical likelihoods that each of the epitopes is presented by one or more of the MHC alleles on the infected cell surface, the set of numerical likelihoods having been identified at least based on received mass spectrometry data; and
 - (c) selecting a subset of the set of epitopes based on the set of numerical likelihoods to generate a set of selected epitopes which are used to generate the MHC class I epitope-encoding nucleic acid sequence.
80. The composition of claim 20, wherein each of the MHC class I epitope-encoding nucleic acid sequences is selected by performing the steps of:
- (a) obtaining at least one of exome, transcriptome, or whole genome infectious disease organism nucleotide sequencing data from the infectious disease organism, wherein the infectious disease organism nucleotide sequencing data is used to obtain data representing peptide sequences of each of a set of epitopes;
 - (b) inputting the peptide sequence of each epitope into a presentation model to generate a set of numerical likelihoods that each of the epitopes is presented by one

or more of the MHC alleles on the infected cell surface, the set of numerical likelihoods having been identified at least based on received mass spectrometry data; and

(c) selecting a subset of the set of epitopes based on the set of numerical likelihoods to generate a set of selected epitopes which are used to generate the at least 20 MHC class I epitope-encoding nucleic acid sequences.

81. The composition of claim 79, wherein a number of the set of selected epitopes is 2-20.
82. The composition of any one of claims 79-81, wherein the presentation model represents dependence between:
 - (a) presence of a pair of a particular one of the MHC alleles and a particular amino acid at a particular position of a peptide sequence; and
 - (b) likelihood of presentation on the infected cell surface, by the particular one of the MHC alleles of the pair, of such a peptide sequence comprising the particular amino acid at the particular position.
83. The composition of any one of claims 79-82, wherein selecting the set of selected epitopes comprises selecting epitopes that have an increased likelihood of being presented on the infected cell surface relative to unselected epitopes based on the presentation model.
84. The composition of any one of claims 79-83, wherein selecting the set of selected epitopes comprises selecting epitopes that have an increased likelihood of being capable of inducing a infectious disease organism-specific immune response in the subject relative to unselected epitopes based on the presentation model.
85. The composition of any one of claims 79-84, wherein selecting the set of selected epitopes comprises selecting epitopes that have an increased likelihood of being capable of being presented to naïve T cells by professional antigen presenting cells (APCs) relative to unselected epitopes based on the presentation model, optionally wherein the APC is a dendritic cell (DC).
86. The composition of any one of claims 79-85, wherein selecting the set of selected epitopes comprises selecting epitopes that have a decreased likelihood of being subject

to inhibition via central or peripheral tolerance relative to unselected epitopes based on the presentation model.

87. The composition of any one of claims 79-86, wherein selecting the set of selected epitopes comprises selecting epitopes that have a decreased likelihood of being capable of inducing an autoimmune response to normal tissue in the subject relative to unselected epitopes based on the presentation model.
88. The composition of any one of claims 79-87, wherein exome or transcriptome nucleotide sequencing data is obtained by performing sequencing on the infected tissue.
89. The composition of claim 88, wherein the sequencing is next generation sequencing (NGS) or any massively parallel sequencing approach.
90. A composition for delivery of a chimpanzee adenovirus (ChAdV)-based expression system, wherein the composition for delivery of the ChAdV-based expression system comprises:

the ChAdV-based expression system, wherein the ChAdV-based expression system comprises a viral particle comprising a ChAdV vector, wherein the ChAdV vector comprises:
 - (a) a ChAdV backbone, wherein the ChAdV backbone comprises:
 - (i) at least one promoter nucleotide sequence, and
 - (ii) at least one polyadenylation (poly(A)) sequence; and
 - (b) a cassette, wherein the cassette comprises:
 - (i) at least one antigen-encoding nucleic acid sequence comprising:
 - a. a nucleic acid sequence encoding an infectious disease organism peptideslected from the group consisting of: a pathogen-derived peptide, a virus-derived peptide, a bacteria-derived peptide, a fungus-derived peptide, and a parasite-derived peptide,
 - b. optionally a 5' linker sequence, and
 - c. optionally a 3' linker sequence; and

wherein the cassette is operably linked to the at least one promoter nucleotide sequence and the at least one poly(A) sequence.

91. A composition for delivery of a ChAdV-based expression system,

wherein the composition for delivery of the ChAdV-based expression system comprises:

the ChAdV-based expression system, wherein the ChAdV-based expression system comprises a viral particle comprising a ChAdV vector, wherein the ChAdV vector comprises:

- (a) a ChAdV backbone, wherein the ChAdV backbone comprises:
 - (i) a modified ChAdV68 sequence comprising at least nucleotides 2 to 36,518 of the sequence set forth in SEQ ID NO:1, wherein the nucleotides 2 to 36,518 lack: (1) nucleotides 577 to 3403 of the sequence shown in SEQ ID NO:1 corresponding to an E1 deletion; (2) nucleotides 27,125 to 31,825 of the sequence shown in SEQ ID NO:1 corresponding to an E3 deletion; and optionally (3) nucleotides 34,916 to 35,642 of the sequence shown in SEQ ID NO:1 corresponding to a partial E4 deletion;
 - (ii) a CMV promoter nucleotide sequence; and
 - (iii) an SV40 polyadenylation (poly(A)) sequence; and
- (b) a cassette, wherein the cassette comprises:
 - (i) at least one antigen-encoding nucleic acid sequence comprising:
 - a. a nucleic acid sequence encoding an infectious disease organism peptide selected from the group consisting of: a pathogen-derived peptide, a virus-derived peptide, a bacteria-derived peptide, a fungus-derived peptide, and a parasite-derived peptide,
 - b. optionally a 5' linker sequence, and
 - c. optionally a 3' linker sequence; and

wherein the cassette is inserted within the E1 deletion and the cassette is operably linked to the CMV promoter nucleotide sequence and the SV40 poly(A) sequence.

92. The composition of claims 90 or 91, wherein the nucleic acid sequence encoding the peptide comprises an epitope-encoding nucleic acid sequence, optionally wherein the nucleic acid sequence encoding the peptide comprises two or more distinct epitope-encoding nucleic acid sequences.
93. The composition of claims 90 or 91, wherein the nucleic acid sequence encoding the peptide comprises between 1-10, between 1-20, between 1-30, between 1-40, between 1-50, between 1-100, between 1-200, between 1-300, between 1-400, or between 1-500 distinct epitope-encoding nucleic acid sequences distinct epitope-encoding nucleic acid sequences, optionally wherein the nucleic acid sequence encoding the peptide comprises between 2-10, between 2-20, between 2-30, between 2-40, between 2-50, between 2-100, between 2-200, between 2-300, between 2-400, or between 2-500 distinct epitope-encoding nucleic acid sequences distinct epitope-encoding nucleic acid sequences.
94. The composition of claims 90 or 91, wherein the nucleic acid sequence encoding the peptide comprises a nucleic acid sequence encoding a peptide selected from: an epitope, a full-length protein, a protein subunit, a protein domain, of the protein expressed in the infectious disease organism, optionally wherein the infectious disease organism comprises two or more distinct nucleic acid sequences encoding a peptide selected from: an epitope, a full-length protein, a protein subunit, a protein domain, and combinations thereof of the protein expressed in the infectious disease organism.
95. The composition of any one of claims claims 90-94, wherein the encoded peptide or peptides is capable of stimulating an immune response when expressed in a subject, optionally wherein the immune response is a T cell response and/or a B cell response.
96. The composition of any one of claims claims 90-95, wherein the infectious disease organism is selected from the group consisting of: Severe acute respiratory syndrome-related coronavirus (SARS), severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), Ebola, HIV, Hepatitis B virus (HBV), influenza, Hepatitis C virus (HCV), Human papillomavirus (HPV), Cytomegalovirus (CMV), Chikungunya virus, Respiratory syncytial virus (RSV), Dengue virus, a orthomyxoviridae family virus, and tuberculosis.

97. The composition of any one of the above claims, wherein the epitope-encoding nucleic acid sequence encodes an epitope known or suspected to be presented by MHC class I on a surface of a cell, optionally wherein the surface of the cell is an infected cell surface, and optionally wherein the cell is a subject's cell.
98. The composition of claim 97, wherein the cell is an infected cell selected from the group consisting of: a pathogen infected cell, a virally infected cell, a bacterially infected cell, a fungally infected cell, and a parasitically infected cell.
99. The composition of claim 98, wherein the virally infected cell is selected from the group consisting of: an HIV infected cell, an HPV infected cell, a SARS infected cell, a SARS-CoV-2 infected cell, an Ebola infected cell, a HBV infected cell, an influenza infected cell, a HCV infected cell, a CMV infected cell, a Chikungunya virus infected cell, a RSV infected cell, a Dengue virus infected cell, a orthomyxoviridae family virus infected cell, and a tuberculosis infected cell.
100. The composition of any one of claims 90, or 92-99, wherein an ordered sequence of each element of the cassette in the composition for delivery of the ChAdV-based expression system is described in the formula, from 5' to 3', comprising

$$P_a-(L5_b-N_c-L3_d)_X-(G5_e-U_f)_Y-G3_g$$

wherein P comprises the at least one promoter sequence operably linked to at least one of the at least one antigen-encoding nucleic acid sequences, where $a = 1$,

N comprises one of the epitope-encoding nucleic acid sequences, wherein the epitope-encoding nucleic acid sequence comprises an MHC class I epitope-encoding nucleic acid sequence, where $c = 1$,

L5 comprises the 5' linker sequence, where $b = 0$ or 1 ,

L3 comprises the 3' linker sequence, where $d = 0$ or 1 ,

G5 comprises one of the at least one nucleic acid sequences encoding a GPGPG amino acid linker, where $e = 0$ or 1 ,

G3 comprises one of the at least one nucleic acid sequences encoding a GPGPG amino acid linker, where $g = 0$ or 1 ,

U comprises one of the at least one MHC class II epitope-encoding nucleic acid sequence, where $f = 1$,

$X = 1$ to 400, where for each X the corresponding N_c is an MHC class I epitope-encoding nucleic acid sequence, and

$Y = 0, 1$, or 2, where for each Y the corresponding U_f is an MHC class II epitope-encoding nucleic acid sequence.

101. The composition of claim 100, wherein for each X the corresponding N_c is a distinct MHC class I epitope-encoding nucleic acid sequence.
102. The composition of claim 100 or 101, wherein for each Y the corresponding U_f is a distinct MHC class II epitope-encoding nucleic acid sequence.
103. The composition of any one of claims 100-102, wherein

$b = 1, d = 1, e = 1, g = 1, h = 1, X = 10, Y = 2$,

P is a CMV promoter sequence,

each N encodes a MHC class I epitope 7-15 amino acids in length,

L5 is a native 5' linker sequence that encodes a native N-terminal amino acid sequence of the MHC I epitope, and wherein the 5' linker sequence encodes a peptide that is at least 3 amino acids in length,

L3 is a native 3' linker sequence that encodes a native C-terminal amino acid sequence of the MHC I epitope, and wherein the 3' linker sequence encodes a peptide that is at least 3 amino acids in length,

U is each of a PADRE class II sequence and a Tetanus toxoid MHC class II sequence, the ChAdV vector comprises a modified ChAdV68 sequence comprising at least nucleotides 2 to 36,518 of the sequence set forth in SEQ ID NO:1, wherein the nucleotides 2 to 36,518 lack: (1) nucleotides 577 to 3403 of the sequence shown in SEQ ID NO:1 corresponding to an E1 deletion; (2) nucleotides 27,125 to 31,825 of the sequence shown in SEQ ID NO:1 corresponding to an E3 deletion; and optionally (3) nucleotides 34,916 to 35,642 of the sequence shown in SEQ ID NO:1 corresponding to a partial E4 deletion, and the antigen cassette is inserted within the E1 deletion, and

each of the MHC class I antigen-encoding nucleic acid sequences encodes a polypeptide that is 25 amino acids in length.

104. The composition of any of the above claims, wherein the composition for delivery of the ChAdV-based expression system is formulated for intramuscular (IM), intradermal (ID), subcutaneous (SC), or intravenous (IV) administration.
105. The composition of any of the above claims, wherein the composition for delivery of the ChAdV-based expression system is formulated for intramuscular (IM) administration.
106. The composition of any one of claims 90, 92-102, or 104-105, wherein the cassette is integrated between the at least one promoter nucleotide sequence and the at least one poly(A) sequence.
107. The composition of any one of claims 90, 92-102, or 104-106, wherein the at least one promoter nucleotide sequence is operably linked to the cassette.
108. The composition of any one of claims 90, 92-102, or 104-107, wherein the ChAdV backbone comprises a ChAdV68 vector backbone.
109. The composition of claim 108, wherein the ChAdV68 vector backbone comprises the sequence set forth in SEQ ID NO:1.
110. The composition of claim 108, wherein the ChAdV68 vector backbone comprises a functional deletion in at least one gene selected from the group consisting of an adenovirus E1A, E1B, E2A, E2B, E3, L1, L2, L3, L4, and L5 gene with reference to a ChAdV68 genome or with reference to the sequence shown in SEQ ID NO:1, optionally wherein the adenoviral backbone or modified ChAdV68 sequence is fully deleted or functionally deleted in: (1) E1A and E1B; or (2) E1A, E1B, and E3 with reference to the adenovirus genome or with reference to the sequence shown in SEQ ID NO:1, optionally wherein the E1 gene is functionally deleted through an E1 deletion of at least nucleotides 577 to 3403 with reference to the sequence shown in SEQ ID NO:1 and optionally wherein the E3 gene is functionally deleted through an E3 deletion of at least nucleotides 27,125 to 31,825 with reference to the sequence shown in SEQ ID NO:1.
111. The composition of claim 108, wherein the ChAdV68 vector backbone comprises one or more deletions between base pair number 577 and 3403 or between base pair 456 and 3014, and optionally wherein the vector further comprises one or more deletions

between base pair 27,125 and 31,825 or between base pair 27,816 and 31,333 of the sequence set forth in SEQ ID NO:1.

112. The composition of claim 108, wherein the ChAdV68 vector backbone comprises at least nucleotides 2 to 36,518 of the sequence set forth in SEQ ID NO:1, wherein the nucleotides 2 to 36,518 lack:

A. nucleotides 577 to 3403 of the sequence shown in SEQ ID NO:1 corresponding to an E1 deletion;

B. nucleotides 27,125 to 31,825 of the sequence shown in SEQ ID NO:1 corresponding to an E3 deletion;

C. nucleotides 34,916 to 35,642 of the sequence shown in SEQ ID NO:1 corresponding to a partial E4 deletion;

D. nucleotides 456 to 3014 of the sequence shown in SEQ ID NO:1;

E. nucleotides 27,816 to 31,333 of the sequence shown in SEQ ID NO:1;

F. nucleotides 3957 to 10346 of the sequence shown in SEQ ID NO:1;

G. nucleotides 21787 to 23370 of the sequence shown in SEQ ID NO:1;

H. nucleotides 33486 to 36193 of the sequence shown in SEQ ID NO:1; or

combinations thereof.

113. The composition of any one of claims 108-112, wherein the ChAdV68 vector backbone comprises a partially deleted E4 gene, optionally wherein the partially deleted E4 gene comprises:

A. the E4 gene sequence shown in SEQ ID NO:1 and that lacks at least nucleotides 34,916 to 35,642 of the sequence shown in SEQ ID NO:1,

B. the E4 gene sequence shown in SEQ ID NO:1 and that lacks at least nucleotides 34,916 to 34,942, nucleotides 34,952 to 35,305 of the sequence shown in SEQ ID NO:1, nucleotides 35,302 to 35,642 of the sequence shown in SEQ ID NO:1, and wherein the vector comprises at least nucleotides 2 to 36,518 of the sequence shown in SEQ ID NO:1,

- C. the E4 gene sequence shown in SEQ ID NO:1 and that lacks at least nucleotides 34,980 to 36,516 of the sequence shown in SEQ ID NO:1, and wherein the vector comprises at least nucleotides 2 to 36,518 of the sequence shown in SEQ ID NO:1,
 - D. the E4 gene sequence shown in SEQ ID NO:1 and that lacks at least nucleotides 34,979 to 35,642 of the sequence shown in SEQ ID NO:1, and wherein the vector comprises at least nucleotides 2 to 36,518 of the sequence shown in SEQ ID NO:1,
 - E. an E4 deletion of at least a partial deletion of E4Orf2, a fully deleted E4Orf3, and at least a partial deletion of E4Orf4,
 - F. an E4 deletion of at least a partial deletion of E4Orf2, at least a partial deletion of E4Orf3, and at least a partial deletion of E4Orf4,
 - G. an E4 deletion of at least a partial deletion of E4Orf1, a fully deleted E4Orf2, and at least a partial deletion of E4Orf3, or
 - H. an E4 deletion of at least a partial deletion of E4Orf2 and at least a partial deletion of E4Orf3.
- 114. The composition of any one of claims 90, 92-102, or 104-113, wherein the wherein the cassette is inserted in the ChAdV backbone at the E1 region, E3 region, and/or any deleted AdV region that allows incorporation of the cassette.
 - 115. The composition of any one of 90, 92-102, or 104-114, wherein the ChAdV backbone is generated from one of a first generation, a second generation, or a helper-dependent adenoviral vector.
 - 116. The composition of any one of claims 90, 92-102, or 104-115, wherein the at least one promoter nucleotide sequence is selected from the group consisting of: a CMV, a SV40, an EF-1, a RSV, a PGK, a HSA, a MCK, and a EBV promoter sequence.
 - 117. The composition of any one of claims 90, 92-102, or 104-115, wherein the at least one promoter nucleotide sequence is a CMV promoter sequence.
 - 118. The composition of any one of the above claims, wherein at least one of the epitope-encoding nucleic acid sequences encodes an epitope that, when expressed and translated, is capable of being presented by MHC class I on a cell of a subject.
 - 119. The composition of any one of the above claims, wherein at least one of the epitope-encoding nucleic acid sequences encodes an epitope that, when expressed and translated, is capable of being presented by MHC class II on a cell of a subject.

120. The composition of any one of claims 90-102, or 104-119, wherein the at least one antigen-encoding nucleic acid sequence comprises two or more antigen-encoding nucleic acid sequences.
121. The composition of claim 120, wherein each antigen-encoding nucleic acid sequence is linked directly to one another.
122. The composition of any one of claims 90-102, or 104-121, wherein each antigen-encoding nucleic acid sequence is linked to a distinct antigen-encoding nucleic acid sequence with a nucleic acid sequence encoding a linker.
123. The composition of claim 122, wherein the linker links two MHC class I epitope-encoding nucleic acid sequences or an MHC class I epitope-encoding nucleic acid sequence to an MHC class II epitope-encoding nucleic acid sequence.
124. The composition of claim 123, wherein the linker is selected from the group consisting of: (1) consecutive glycine residues, at least 2, 3, 4, 5, 6, 7, 8, 9, or 10 residues in length; (2) consecutive alanine residues, at least 2, 3, 4, 5, 6, 7, 8, 9, or 10 residues in length; (3) two arginine residues (RR); (4) alanine, alanine, tyrosine (AAY); (5) a consensus sequence at least 2, 3, 4, 5, 6, 7, 8, 9, or 10 amino acid residues in length that is processed efficiently by a mammalian proteasome; and (6) one or more native sequences flanking the antigen derived from the cognate protein of origin and that is at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, or 2-20 amino acid residues in length.
125. The composition of claim 122, wherein the linker links two MHC class II epitope-encoding nucleic acid sequences or an MHC class II sequence to an MHC class I epitope-encoding nucleic acid sequence.
126. The composition of claim 125, wherein the linker comprises the sequence GPGPG.
127. The composition of any one of claims 90-102, or 104-126, wherein the antigen-encoding nucleic acid sequences is linked, operably or directly, to a separate or contiguous sequence that enhances the expression, stability, cell trafficking, processing and presentation, and/or immunogenicity of the antigen-encoding nucleic acid sequence.
128. The composition of claim 127, wherein the separate or contiguous sequence comprises at least one of: a ubiquitin sequence, a ubiquitin sequence modified to increase proteasome targeting (e.g., the ubiquitin sequence contains a Gly to Ala substitution at position 76), an immunoglobulin signal sequence (e.g., IgK), a major histocompatibility

class I sequence, lysosomal-associated membrane protein (LAMP)-1, human dendritic cell lysosomal-associated membrane protein, and a major histocompatibility class II sequence; optionally wherein the ubiquitin sequence modified to increase proteasome targeting is A76.

129. The composition of any of the above claims, wherein the epitope-encoding nucleic acid sequence comprises at least one alteration that makes the encoded epitope have increased binding affinity to its corresponding MHC allele relative to the translated, corresponding wild-type nucleic acid sequence.
130. The composition of any of the above claims, wherein the epitope-encoding nucleic acid sequence comprises at least one alteration that makes the encoded epitope have increased binding stability to its corresponding MHC allele relative to the translated, corresponding wild-type nucleic acid sequence.
131. The composition of any of the above claims, wherein the epitope-encoding nucleic acid sequence comprises at least one alteration that makes the encoded epitope have an increased likelihood of presentation on its corresponding MHC allele relative to the translated, corresponding wild-type nucleic acid sequence.
132. The composition of any of the above claims, wherein the at least one alteration comprises a point mutation, a frameshift mutation, a non-frameshift mutation, a deletion mutation, an insertion mutation, a splice variant, a genomic rearrangement, or a proteasome-generated spliced antigen.
133. The composition of any of the above claims, wherein the epitope-encoding nucleic acid sequence encodes an epitope known or suspected to be expressed in a subject known or suspected to have an infection.
134. The composition of claim 133, wherein the infection is selected from the group consisting of: a pathogen infection, a viral infection, a bacterial infection, an fungal infection, and a parasitic infection.
135. The composition of claim 134, wherein the viral infection is selected from the group consisting of: an HIV infection, an HPV infection, a SARS infection, a SARS-CoV-2 infection, an Ebola infection, a HBV infection, an influenza infection, a HCV infection, a CMV infection, a Chikungunya virus infection, a RSV infection, a Dengue virus infection, a orthomyxoviridae family virus infection, and a tuberculosis infection.
136. The composition of claim 134, wherein the bacterial infection is a tuberculosis infection.

137. The composition of any one of claims 90-102, or 104-136, wherein the at least one antigen-encoding nucleic acid sequence comprises at least 2-10, 2, 3, 4, 5, 6, 7, 8, 9, or 10 antigen-encoding nucleic acid sequences, optionally wherein each antigen-encoding nucleic acid sequence encodes a distinct antigen-encoding nucleic acid sequence.
138. The composition of any one of claims 90-102, or 104-136, wherein the at least one antigen-encoding nucleic acid sequence comprises at least 11-20, 15-20, 11-100, 11-200, 11-300, 11-400, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 or up to 400 antigen-encoding nucleic acid sequences, optionally wherein each antigen-encoding nucleic acid sequence encodes a distinct antigen-encoding nucleic acid sequence.
139. The composition any one of claims 90-102, or 104-136, wherein the at least one antigen-encoding nucleic acid sequence comprises at least 11-20, 15-20, 11-100, 11-200, 11-300, 11-400, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 or up to 400 antigen-encoding nucleic acid sequences.
140. The composition of any one of claims 90-102, or 104-136, wherein the at least one antigen-encoding nucleic acid sequence comprises at least 2-400 antigen-encoding nucleic acid sequences and wherein at least two of the antigen-encoding nucleic acid sequences encode epitope sequences or portions thereof that are presented by MHC class I on a cell surface.
141. The composition of 103, wherein at least two of the MHC class I epitopes are presented by MHC class I on the infected cell surface.
142. The composition of any one of claims 90-102, or 104-141, wherein the epitope-encoding nucleic acid sequences comprises at least one MHC class I epitope-encoding nucleic acid sequence, and wherein each antigen-encoding nucleic acid sequence encodes a polypeptide sequence between 8 and 35 amino acids in length, optionally 9-17, 9-25, 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 or 35 amino acids in length.
143. The composition of any one of claims 100-102, or 104-142, wherein the at least one MHC class II epitope-encoding nucleic acid sequence is present.
144. The composition of any one of claims 100-102, or 104-142, wherein the at least one MHC class II epitope-encoding nucleic acid sequence is present and comprises at least one MHC class II epitope-encoding nucleic acid sequence that comprises at least one alteration that makes the encoded epitope sequence distinct from the corresponding peptide sequence encoded by a wild-type nucleic acid sequence.

145. The composition of any one of claims 90-102, or 104-144, wherein the epitope-encoding nucleic acid sequence comprises an MHC class II epitope-encoding nucleic acid sequence and wherein each antigen-encoding nucleic acid sequence encodes a polypeptide sequence that is 12-20, 12, 13, 14, 15, 16, 17, 18, 19, 20, or 20-40 amino acids in length.
146. The composition of any one of claims 90-102, or 104-145, wherein the epitope-encoding nucleic acid sequences comprises an MHC class II epitope-encoding nucleic acid sequence, wherein the at least one MHC class II epitope-encoding nucleic acid sequence is present, and wherein the at least one MHC class II epitope-encoding nucleic acid sequence comprises at least one universal MHC class II epitope-encoding nucleic acid sequence, optionally wherein the at least one universal sequence comprises at least one of Tetanus toxoid and PADRE.
147. The composition of any one of claims 90, 92-102, or 104-146, wherein the at least one promoter nucleotide sequence is inducible.
148. The composition of any one of claims 90, 92-102, or 104-146, wherein the at least one promoter nucleotide sequence is non-inducible.
149. The composition of any one of claims 90, 92-102, or 104-148, wherein the at least one poly(A) sequence comprises a Bovine Growth Hormone (BGH) SV40 polyA sequence.
150. The composition of any one claims 90, 92-102, or 104-149, wherein the at least one poly(A) sequence is at least 20 , at least 30, at least 40, at least 50, at least 60, at least 70, at least 80, or at least 90 consecutive A nucleotides.
151. The composition of any one of claims 90, 92-102, or 104-149, wherein the at least one poly(A) sequence is at least 100 consecutive A nucleotides.
152. The composition of any of the above claims, wherein the cassette further comprises at least one of: an intron sequence, a woodchuck hepatitis virus posttranscriptional regulatory element (WPRE) sequence, an internal ribosome entry sequence (IRES) sequence, a nucleotide sequence encoding a 2A self cleaving peptide sequence, a nucleotide sequence encoding a Furin cleavage site, or a sequence in the 5' or 3' non-coding region known to enhance the nuclear export, stability, or translation efficiency of mRNA that is operably linked to at least one of the at least one antigen-encoding nucleic acid sequences.
153. The composition of any of the above claims, wherein the cassette further comprises a reporter gene, including but not limited to, green fluorescent protein (GFP), a GFP

variant, secreted alkaline phosphatase, luciferase, a luciferase variant, or a detectable peptide or epitope.

154. The composition of claim 153, wherein the detectable peptide or epitope is selected from the group consisting of an HA tag, a Flag tag, a His-tag, or a V5 tag.
155. The composition of any of the above claims, wherein the one or more vectors further comprises one or more nucleic acid sequences encoding at least one immune modulator.
156. The composition of claim 155, wherein the immune modulator is an anti-CTLA4 antibody or an antigen-binding fragment thereof, an anti-PD-1 antibody or an antigen-binding fragment thereof, an anti-PD-L1 antibody or an antigen-binding fragment thereof, an anti-4-1BB antibody or an antigen-binding fragment thereof, or an anti-OX-40 antibody or an antigen-binding fragment thereof.
157. The composition of claim 156, wherein the antibody or antigen-binding fragment thereof is a Fab fragment, a Fab' fragment, a single chain Fv (scFv), a single domain antibody (sdAb) either as single specific or multiple specificities linked together (e.g., camelid antibody domains), or full-length single-chain antibody (e.g., full-length IgG with heavy and light chains linked by a flexible linker).
158. The composition of claim 156 or 157, wherein the heavy and light chain sequences of the antibody are a contiguous sequence separated by either a self-cleaving sequence such as 2A or IRES; or the heavy and light chain sequences of the antibody are linked by a flexible linker such as consecutive glycine residues.
159. The composition of claim 155, wherein the immune modulator is a cytokine.
160. The composition of claim 159, wherein the cytokine is at least one of IL-2, IL-7, IL-12, IL-15, or IL-21 or variants thereof of each.
161. The composition of any one of claims 90-102, or 104-160, wherein the epitope-encoding nucleic acid sequence comprises a MHC class I epitope-encoding nucleic acid sequence, and wherein the MHC class I epitope-encoding nucleic acid sequence is selected by performing the steps of:
 - (a) obtaining at least one of exome, transcriptome, or whole genome infectious disease organism nucleotide sequencing data from the infectious disease organism, wherein the infectious disease organism nucleotide sequencing data is used to obtain data representing peptide sequences of each of a set of epitopes;

- (b) inputting the peptide sequence of each epitope into a presentation model to generate a set of numerical likelihoods that each of the epitopes is presented by one or more of the MHC alleles on the infected cell surface, the set of numerical likelihoods having been identified at least based on received mass spectrometry data; and
 - (c) selecting a subset of the set of epitopes based on the set of numerical likelihoods to generate a set of selected epitopes which are used to generate the MHC class I epitope-encoding nucleic acid sequence.
162. The composition of claim 103, wherein each of the MHC class I epitope-encoding nucleic acid sequences is selected by performing the steps of:
- (a) obtaining at least one of exome, transcriptome, or whole genome infectious disease organism nucleotide sequencing data from the infectious disease organism, wherein the infectious disease organism nucleotide sequencing data is used to obtain data representing peptide sequences of each of a set of epitopes;
 - (b) inputting the peptide sequence of each epitope into a presentation model to generate a set of numerical likelihoods that each of the epitopes is presented by one or more of the MHC alleles on the infected cell surface, the set of numerical likelihoods having been identified at least based on received mass spectrometry data; and
 - (c) selecting a subset of the set of epitopes based on the set of numerical likelihoods to generate a set of selected epitopes which are used to generate the at least 20 MHC class I epitope-encoding nucleic acid sequences.
163. The composition of claim 161, wherein a number of the set of selected epitopes is 2-20.
164. The composition of claim 161-163, wherein the presentation model represents dependence between:
- (a) presence of a pair of a particular one of the MHC alleles and a particular amino acid at a particular position of a peptide sequence; and
 - (b) likelihood of presentation on the infected cell surface, by the particular one of the MHC alleles of the pair, of such a peptide sequence comprising the particular amino acid at the particular position.

165. The composition of claim 161-164, wherein selecting the set of selected epitopes comprises selecting epitopes that have an increased likelihood of being presented on the infected cell surface relative to unselected epitopes based on the presentation model.
166. The composition of claim 161-165, wherein selecting the set of selected epitopes comprises selecting epitopes that have an increased likelihood of being capable of inducing a infectious disease organism-specific immune response in the subject relative to unselected epitopes based on the presentation model.
167. The composition of claim 161-166, wherein selecting the set of selected epitopes comprises selecting epitopes that have an increased likelihood of being capable of being presented to naïve T cells by professional antigen presenting cells (APCs) relative to unselected epitopes based on the presentation model, optionally wherein the APC is a dendritic cell (DC).
168. The composition of claim 161-167, wherein selecting the set of selected epitopes comprises selecting epitopes that have a decreased likelihood of being subject to inhibition via central or peripheral tolerance relative to unselected epitopes based on the presentation model.
169. The composition of claim 161-168, wherein selecting the set of selected epitopes comprises selecting epitopes that have a decreased likelihood of being capable of inducing an autoimmune response to normal tissue in the subject relative to unselected epitopes based on the presentation model.
170. The composition of claim 161-169, wherein exome or transcriptome nucleotide sequencing data is obtained by performing sequencing on the infected tissue.
171. The composition of claim 170, wherein the sequencing is next generation sequencing (NGS) or any massively parallel sequencing approach.
172. The composition of any of the above claims, wherein the cassette comprises junctional epitope sequences formed by adjacent sequences in the cassette.
173. The composition of claim 172, wherein at least one or each junctional epitope sequence has an affinity of greater than 500 nM for MHC.
174. The composition of claims 172 or 173, wherein each junctional epitope sequence is non-self.
175. The composition of any of the above claims, wherein the cassette does not encode a non-therapeutic MHC class I or class II epitope nucleic acid sequence comprising a

- translated, wild-type nucleic acid sequence, wherein the non-therapeutic epitope is predicted to be displayed on an MHC allele of the subject.
176. The composition of claim 175, wherein the non-therapeutic predicted MHC class I or class II epitope sequence is a junctional epitope sequence formed by adjacent sequences in the cassette.
177. The composition of claims 172-176, wherein the prediction is based on presentation likelihoods generated by inputting sequences of the non-therapeutic epitopes into a presentation model.
178. The composition of any one of claims 172-177, wherein an order of the antigen-encoding nucleic acid sequences in the cassette is determined by a series of steps comprising:
- (a) generating a set of candidate cassette sequences corresponding to different orders of the antigen-encoding nucleic acid sequences;
 - (b) determining, for each candidate cassette sequence, a presentation score based on presentation of non-therapeutic epitopes in the candidate cassette sequence; and
 - (c) selecting a candidate cassette sequence associated with a presentation score below a predetermined threshold as the cassette sequence for a vaccine.
179. The composition of any of the above claims, wherein the composition for delivery of the ChAdV-based expression system is formulated in a pharmaceutical composition comprising a pharmaceutically acceptable carrier.
180. A kit comprising the composition for delivery of the ChAdV-based expression system of any of the above composition claims, and instructions for use.
181. The composition of any of the above claims, wherein one or more of the epitope-encoding nucleic acid sequences are derived from an infection in or an infected cell of a subject.
182. The composition of any of the above claims, wherein each of the epitope-encoding nucleic acid sequences are derived from an infection in or an infected cell of a subject.
183. The composition of any of the above claims, wherein one or more of the epitope-encoding nucleic acid sequences are not derived from an infection in or an infected cell of a subject.
184. The composition of any of the above claims, wherein each of the epitope-encoding nucleic acid sequences are not derived from an infection in or an infected cell of a subject.

185. A composition for delivery of a self-replicating alphavirus-based expression system,

wherein the composition for delivery of the self-replicating alphavirus-based expression system comprises:

(A) the self-replicating alphavirus-based expression system, wherein the self-replicating alphavirus-based expression system comprises one or more vectors, wherein the one or more vectors comprises:

(a) an RNA alphavirus backbone, wherein the RNA alphavirus backbone comprises:

(i) at least one promoter nucleotide sequence, and

(ii) at least one polyadenylation (poly(A)) sequence; and

(b) a cassette, wherein the cassette comprises:

(i) at least one antigen-encoding nucleic acid sequence comprising:

a. a nucleic acid sequence encoding an infectious disease organism peptide selected from the group consisting of: a pathogen-derived peptide, a virus-derived peptide, a bacteria-derived peptide, a fungus-derived peptide, and a parasite-derived peptide, wherein the infectious disease organism is selected from the group consisting of: Severe acute respiratory syndrome-related coronavirus (SARS), severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), Ebola, HIV, Hepatitis B virus (HBV), influenza, Hepatitis C virus (HCV), Human papillomavirus (HPV), Cytomegalovirus (CMV), Chikungunya virus, Respiratory syncytial virus (RSV), Dengue virus, a orthomyxoviridae family virus, and tuberculosis;

b. optionally a 5' linker sequence, and

c. optionally a 3' linker sequence;

(ii) optionally, a second promoter nucleotide sequence operably linked to the at least one antigen-encoding nucleic acid sequence; and

(iii) optionally, at least one second poly(A) sequence, wherein the second poly(A) sequence is a native poly(A) sequence or an exogenous poly(A) sequence to the alphavirus, and

(B) a lipid-nanoparticle (LNP), wherein the LNP encapsulates the self-replicating alphavirus-based expression system.

186. A composition for delivery of a chimpanzee adenovirus (ChAdV)-based expression system, wherein the composition for delivery of the ChAdV-based expression system comprises:

the ChAdV-based expression system, wherein the ChAdV-based expression system comprises a viral particle comprising a ChAdV vector, wherein the ChAdV vector comprises:

(a) a ChAdV backbone, wherein the ChAdV backbone comprises:

(i) at least one promoter nucleotide sequence, and

(ii) at least one polyadenylation (poly(A)) sequence; and

(b) a cassette, wherein the cassette comprises:

(i) at least one antigen-encoding nucleic acid sequence comprising:

a. a nucleic acid sequence encoding an infectious disease organism peptide selected from the group consisting of: a pathogen-derived peptide, a virus-derived peptide, a bacteria-derived peptide, a fungus-derived peptide, and a parasite-derived peptide, wherein the infectious disease organism is selected from the group consisting of: Severe acute respiratory syndrome-related coronavirus (SARS), severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), Ebola, HIV, Hepatitis B virus (HBV), influenza, Hepatitis C virus (HCV), Human papillomavirus (HPV), Cytomegalovirus (CMV), Chikungunya virus, Respiratory syncytial virus (RSV), Dengue virus, a orthomyxoviridae family virus, and tuberculosis;

b. optionally a 5' linker sequence, and

c. optionally a 3' linker sequence; and

wherein the cassette is operably linked to the at least one promoter nucleotide sequence and the at least one poly(A) sequence.

187. The composition of claim 185 or 186, wherein the self-replicating alphavirus-based expression system or the ChAdV-based expression system comprise any of the features of any one of the above claims.
188. A method for stimulating an immune response in a subject, the method comprising administering to the subject a composition for delivery of a self-replicating alphavirus-based expression system and/or administering to the subject a composition for delivery of a chimpanzee adenovirus (ChAdV)-based expression system, and wherein either:
- a. the composition for delivery of the ChAdV-based expression system comprises the ChAdV-based expression system,
 - b. wherein the composition for delivery of the self-replicating alphavirus-based expression system comprises the self-replicating alphavirus-based expression system, or
 - c. the composition for delivery of the ChAdV-based expression system comprises the ChAdV-based expression system, and wherein the composition for delivery of the self-replicating alphavirus-based expression system comprises the self-replicating alphavirus-based expression system.
189. The method of claim 188, wherein the composition for delivery of the ChAdV-based expression system is administered as a priming dose and either the composition for delivery of the ChAdV-based expression system or the composition for delivery of the self-replicating alphavirus-based expression system is administered as one or more boosting doses.
190. The method of claim 188, wherein the composition for delivery of the self-replicating alphavirus-based expression system is administered as a priming dose either the composition for delivery of the ChAdV-based expression system or the composition for delivery of the self-replicating alphavirus-based expression system is administered as one or more boosting doses.
191. The method of any of claims 189-190, wherein two or more boosting doses are administered.

192. The method of any of claims 189-190, wherein 1, 2, 3, 4, 5, 6, 7, or 8 boosting doses are administered.
193. The method of any one of the above method claims, wherein the composition for delivery of the ChAdV-based expression system is administered intramuscularly (IM), intradermally (ID), subcutaneously (SC), or intravenously (IV).
194. The method of any one of the above method claims, wherein the composition for delivery of the ChAdV-based expression system is administered (IM).
195. The method of claim 194, wherein the IM administration is administered at separate injection sites.
196. The method of claim 195, wherein the separate injection sites are in opposing deltoid muscles.
197. The method of claim 195, wherein the separate injection sites are in gluteus or rectus femoris sites on each side.
198. The method of any one of the above method claims, wherein the composition for delivery of the self-replicating alphavirus-based expression system is administered intramuscularly (IM), intradermally (ID), subcutaneously (SC), or intravenously (IV).
199. The method of any one of the above method claims, wherein the composition for delivery of the self-replicating alphavirus-based expression system is administered (IM).
200. The method of claim 199, wherein the IM administration is administered at separate injection sites.
201. The method of claim 200, wherein the separate injection sites are in opposing deltoid muscles.
202. The method of claim 200, wherein the separate injection sites are in gluteus or rectus femoris sites on each side.
203. The method of any one of claims 199-202, wherein the injection site of the one or more boosting doses is as close as possible to the injection site of the priming dose.
204. The method of any one of the above method claims, further comprising determining or having determined the HLA-haplotype of the subject.
205. The method of any one of the above method claims, wherein the composition for delivery of the self-replicating alphavirus-based expression system comprises:

(A) the self-replicating alphavirus-based expression system, wherein the self-replicating alphavirus-based expression system comprises one or more vectors, wherein the one or more vectors comprises:

- (a) an RNA alphavirus backbone, wherein the RNA alphavirus backbone comprises:
 - (i) at least one promoter nucleotide sequence, and
 - (ii) at least one polyadenylation (poly(A)) sequence; and
- (b) a cassette, wherein the cassette comprises:
 - (i) at least one antigen-encoding nucleic acid sequence comprising:
 - a. a nucleic acid sequence encoding an infectious disease organism peptide selected from the group consisting of: a pathogen-derived peptide, a virus-derived peptide, a bacteria-derived peptide, a fungus-derived peptide, and a parasite-derived peptide,
 - b. optionally a 5' linker sequence, and
 - c. optionally a 3' linker sequence;
 - (ii) optionally, a second promoter nucleotide sequence operably linked to the at least one antigen-encoding nucleic acid sequence; and
 - (iii) optionally, at least one second poly(A) sequence, wherein the second poly(A) sequence is a native poly(A) sequence or an exogenous poly(A) sequence to the alphavirus, and

(B) a lipid-nanoparticle (LNP), wherein the LNP encapsulates the self-replicating alphavirus-based expression system.

206. The method of any one of the above method claims, wherein the composition for delivery of the self-replicating alphavirus-based expression system comprises,

(A) the self-replicating alphavirus-based expression system, wherein the self-replicating alphavirus-based expression system comprises one or more vectors, wherein the one or more vectors comprises:

(a) an RNA alphavirus backbone, wherein the RNA alphavirus backbone comprises the nucleic acid sequence set forth in SEQ ID NO:6, wherein the RNA alphavirus backbone sequence comprises a 26S promoter nucleotide sequence and a poly(A) sequence, wherein the 26S promoter sequence is endogenous to the RNA alphavirus backbone, and wherein the poly(A) sequence is endogenous to the RNA alphavirus backbone; and

(b) a cassette integrated between the 26S promoter nucleotide sequence and the poly(A) sequence, wherein the cassette is operably linked to the 26S promoter nucleotide sequence, and wherein the cassette comprises at least one antigen-encoding nucleic acid sequence comprising:

- a. a nucleic acid sequence encoding an infectious disease organism peptide selected from the group consisting of: a pathogen-derived peptide, a virus-derived peptide, a bacteria-derived peptide, a fungus-derived peptide, and a parasite-derived peptide,
- b. optionally a 5' linker sequence, and
- c. optionally a 3' linker sequence; and

(B) a lipid-nanoparticle (LNP), wherein the LNP encapsulates the self-replicating alphavirus-based expression system.

207. The method of any one of the above method claims 188-206, wherein the nucleic acid sequence encoding the peptide expressed in the infectious disease organism comprises an epitope-encoding nucleic acid sequence.
208. The method of any one of the above method claims 188-206, wherein the nucleic acid sequence encoding the peptide expressed in the infectious disease organism comprises two or more distinct epitope-encoding nucleic acid sequences.
209. The method of any one of the above method claims 188-206, wherein the nucleic acid sequence encoding the peptide expressed in the infectious disease organism comprises two or more distinct epitope-encoding nucleic acid sequences.
210. The method of any one of the above method claims 188-206, wherein the nucleic acid sequence encoding the peptide expressed in the infectious disease organism comprises between 1-10, between 1-20, between 1-30, between 1-40, between 1-50, between 1-100, between 1-200, between 1-300, between 1-400, or between 1-500 distinct epitope-encoding nucleic acid sequences distinct epitope-encoding nucleic acid sequences.

211. The method of any one of the above method claims 188-206, wherein the nucleic acid sequence encoding the peptide expressed in the infectious disease organism comprises between 2-10, between 2-20, between 2-30, between 2-40, between 2-50, between 2-100, between 2-200, between 2-300, between 2-400, or between 2-500 distinct epitope-encoding nucleic acid sequences distinct epitope-encoding nucleic acid sequences.
212. The method of any one of the above method claims 188-206, wherein the nucleic acid sequence encoding the peptide expressed in the infectious disease organism comprises a nucleic acid sequence encoding a full-length protein expressed in the infectious disease organism.
213. The method of any one of the above method claims 188-206, wherein the nucleic acid sequence encoding the peptide expressed in the infectious disease organism comprises a nucleic acid sequence encoding a protein domain of the protein expressed in the infectious disease organism.
214. The method of any one of the above method claims 188-206, wherein the nucleic acid sequence encoding the peptide expressed in the infectious disease organism comprises a nucleic acid sequence encoding a protein subunit of the protein expressed in the infectious disease organism.
215. The method of any one of the above method claims 188-206, wherein the nucleic acid sequence encoding the peptide expressed in the infectious disease organism comprises two or more distinct nucleic acid sequences encoding a peptide selected from: an epitope, a full-length protein, a protein subunit, a protein domain, and combinations thereof of the protein expressed in the infectious disease organism.
216. The method of any one of the above method claims 188-215, wherein the encoded peptide or peptides is capable of stimulating an immune response when expressed in a subject.
217. The method of any one of the above method claims 188-215, wherein the encoded peptide or peptides is capable of stimulating a T cell response when expressed in a subject.
218. The method of any one of the above method claims 188-215, wherein the encoded peptide or peptides is capable of stimulating a B cell response when expressed in a subject.

219. The method of any one of the above method claims 188-215, wherein the encoded peptide or peptides is capable of stimulating a T cell response and a B cell response when expressed in a subject.
220. The method of any one of the above method claims 188-219, wherein the infectious disease organism is selected from the group consisting of: Severe acute respiratory syndrome-related coronavirus (SARS), severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), Ebola, HIV, Hepatitis B virus (HBV), influenza, Hepatitis C virus (HCV), Human papillomavirus (HPV), Cytomegalovirus (CMV), Chikungunya virus, Respiratory syncytial virus (RSV), Dengue virus, a orthomyxoviridae family virus, and tuberculosis.
221. The method of any one of the above method claims, or 207-220, wherein an ordered sequence of each element of the cassette in the composition for delivery of the self-replicating alphavirus-based expression system is described in the formula, from 5' to 3', comprising

$$P_a-(L5_b-N_c-L3_d)_x-(G5_e-U_f)_y-G3_g$$

wherein P comprises the second promoter nucleotide sequence, where a = 0 or 1,

N comprises one of the epitope-encoding nucleic acid sequences, wherein the epitope-encoding nucleic acid sequence comprises an MHC class I epitope-encoding nucleic acid sequence, where c = 1,

L5 comprises the 5' linker sequence, where b = 0 or 1,

L3 comprises the 3' linker sequence, where d = 0 or 1,

G5 comprises one of the at least one nucleic acid sequences encoding a GPGPG amino acid linker, where e = 0 or 1,

G3 comprises one of the at least one nucleic acid sequences encoding a GPGPG amino acid linker, where g = 0 or 1,

U comprises one of the at least one MHC class II epitope-encoding nucleic acid sequence, where f = 1,

X = 1 to 400, where for each X the corresponding N_c is an MHC class I epitope-encoding nucleic acid sequence, and

Y = 0, 1, or 2, where for each Y the corresponding U_f is an MHC class II epitope-encoding nucleic acid sequence.

- 222. The method of claim 221, wherein for each X the corresponding N_c is a distinct MHC class I epitope-encoding nucleic acid sequence.
- 223. The method of claim 221 or 222, wherein for each Y the corresponding U_f is a distinct MHC class II epitope-encoding nucleic acid sequence.
- 224. The method of any one of the above method claims 221-223, wherein

a = 0, b = 1, d = 1, e = 1, g = 1, h = 1, X = 20, Y = 2,

the at least one promoter nucleotide sequence is a single 26S promoter nucleotide sequence provided by the RNA alphavirus backbone,

the at least one polyadenylation poly(A) sequence is a poly(A) sequence of at least 100 consecutive A nucleotides provided by the RNA alphavirus backbone,

the cassette is integrated between the 26S promoter nucleotide sequence and the poly(A) sequence, wherein the cassette is operably linked to the 26S promoter nucleotide sequence and the poly(A) sequence,

each N encodes a MHC class I epitope 7-15 amino acids in length,

L5 is a native 5' linker sequence that encodes a native N-terminal amino acid sequence of the MHC I epitope, and wherein the 5' linker sequence encodes a peptide that is at least 3 amino acids in length,

L3 is a native 3' linker sequence that encodes a native C-terminal amino acid sequence of the MHC I epitope, and wherein the 3' linker sequence encodes a peptide that is at least 3 amino acids in length,

U is each of a PADRE class II sequence and a Tetanus toxoid MHC class II sequence,

the RNA alphavirus backbone is the sequence set forth in SEQ ID NO:6, and

each of the MHC class I epitope-encoding nucleic acid sequences encodes a polypeptide that is between 13 and 25 amino acids in length.

- 225. The method of any of the above claims, wherein the LNP comprises a lipid selected from the group consisting of: an ionizable amino lipid, a phosphatidylcholine, cholesterol, a PEG-based coat lipid, or a combination thereof.
- 226. The method of any of the above claims, wherein the LNP comprises an ionizable amino lipid, a phosphatidylcholine, cholesterol, and a PEG-based coat lipid.
- 227. The method of claim 225 or 226, wherein the ionizable amino lipids comprise MC3-like (dilinoleylmethyl-4-dimethylaminobutyrate) molecules.
- 228. The method of any of the above claims, wherein the LNP-encapsulated expression system has a diameter of about 100nm.
- 229. The method of any one of the above method claims 205, 207-223, or 225-228, wherein the cassette is integrated between the at least one promoter nucleotide sequence and the at least one poly(A) sequence.
- 230. The method of any one of the above method claims 205, 207-223, or 225-229, wherein the at least one promoter nucleotide sequence is operably linked to the cassette.
- 231. The method of any one of the above method claims 205, 207-223, or 225-230, wherein the one or more vectors comprise one or more +-stranded RNA vectors.
- 232. The method of claim 231 wherein the one or more +-stranded RNA vectors comprise a 5' 7-methylguanosine (m7g) cap.
- 233. The method of claim 231 or 232, wherein the one or more +-stranded RNA vectors are produced by *in vitro* transcription.
- 234. The method of any one of the above method claims 205, 207-223, or 225-233, wherein the one or more vectors are self-replicating within a mammalian cell.
- 235. The method of any one of the above method 205, 207-223, or 225-234, wherein the RNA alphavirus backbone comprises at least one nucleotide sequence of an Aura virus, a Fort Morgan virus, a Venezuelan equine encephalitis virus, a Ross River virus, a Semliki Forest virus, a Sindbis virus, or a Mayaro virus.
- 236. The method of any one of the above method claims 205, 207-223, or 225-234, wherein the RNA alphavirus backbone comprises at least one nucleotide sequence of a Venezuelan equine encephalitis virus.

237. The method of claim 235 or 236, wherein the RNA alphavirus backbone comprises at least sequences for nonstructural protein-mediated amplification, a 26S promoter sequence, a poly(A) sequence, a nonstructural protein 1 (nsP1) gene, a nsP2 gene, a nsP3 gene, and a nsP4 gene encoded by the nucleotide sequence of the Aura virus, the Fort Morgan virus, the Venezuelan equine encephalitis virus, the Ross River virus, the Semliki Forest virus, the Sindbis virus, or the Mayaro virus.
238. The method of claim 235 or 236, wherein the RNA alphavirus backbone comprises at least sequences for nonstructural protein-mediated amplification, a 26S promoter sequence, and a poly(A) sequence encoded by the nucleotide sequence of the Aura virus, the Fort Morgan virus, the Venezuelan equine encephalitis virus, the Ross River virus, the Semliki Forest virus, the Sindbis virus, or the Mayaro virus.
239. The method of claim 237 or 238, wherein sequences for nonstructural protein-mediated amplification are selected from the group consisting of: an alphavirus 5' UTR, a 51-nt CSE, a 24-nt CSE, a 26S subgenomic promoter sequence, a 19-nt CSE, an alphavirus 3' UTR, or combinations thereof.
240. The method of any one of the above method claims 237-239, wherein the RNA alphavirus backbone does not encode structural virion proteins capsid, E2 and E1.
241. The method of claim 240, wherein the cassette is inserted in place of structural virion proteins within the nucleotide sequence of the Aura virus, the Fort Morgan virus, the Venezuelan equine encephalitis virus, the Ross River virus, the Semliki Forest virus, the Sindbis virus, or the Mayaro virus.
242. The method of claim 235 or 236, wherein the Venezuelan equine encephalitis virus comprises the sequence of SEQ ID NO:3 or SEQ ID NO:5.
243. The method of claim 235 or 236, wherein the Venezuelan equine encephalitis virus comprises the sequence of SEQ ID NO:3 or SEQ ID NO:5 further comprising a deletion between base pair 7544 and 11175.
244. The method of claim 243, wherein the RNA alphavirus backbone comprises the sequence set forth in SEQ ID NO:6 or SEQ ID NO:7.
245. The method of claim 243 or 244, wherein the cassette is inserted at position 7544 to replace the deletion between base pairs 7544 and 11175 as set forth in the sequence of SEQ ID NO:3 or SEQ ID NO:5.
246. The method of claim 241-245, wherein the insertion of the cassette provides for transcription of a polycistronic RNA comprising the nsP1-4 genes and the at least one

nucleic acid sequence, wherein the nsP1-4 genes and the at least one nucleic acid sequence are in separate open reading frames.

247. The method of any one of the above method claims 205, 207-223, or 225-246, wherein the at least one promoter nucleotide sequence is the native 26S promoter nucleotide sequence encoded by the RNA alphavirus backbone.
248. The method of any one of the above method claims 205, 207-223, or 225-246, wherein the at least one promoter nucleotide sequence is an exogenous RNA promoter.
249. The method of any one of the above method claims 205, 207-223, or 225-248, wherein the second promoter nucleotide sequence is a 26S promoter nucleotide sequence.
250. The method of any one of the above method claims 205, 207-223, or 225-248, wherein the second promoter nucleotide sequence comprises multiple 26S promoter nucleotide sequences, wherein each 26S promoter nucleotide sequence provides for transcription of one or more of the separate open reading frames.
251. The method of any one of the above method claims, wherein the one or more vectors are each at least 300nt in size.
252. The method of any one of the above method claims, wherein the one or more vectors are each at least 1kb in size.
253. The method of any one of the above method claims, wherein the one or more vectors are each 2kb in size.
254. The method of any one of the above method claims, wherein the one or more vectors are each less than 5kb in size.
255. The method of any one of the above method claims 205-223, or 225-254, wherein the at least one antigen-encoding nucleic acid sequence comprises two or more antigen-encoding nucleic acid sequences.
256. The method of claim 255, wherein each antigen-encoding nucleic acid sequence is linked directly to one another.
257. The method of any one of the above method claims 205-223, or 225-256, wherein each antigen-encoding nucleic acid sequence is linked to a distinct antigen-encoding nucleic acid sequence with a nucleic acid sequence encoding a linker.
258. The method of claim 257, wherein the linker links two MHC class I epitope-encoding nucleic acid sequences or an MHC class I epitope-encoding nucleic acid sequence to an MHC class II epitope-encoding nucleic acid sequence.

259. The method of claim 258, wherein the linker is selected from the group consisting of: (1) consecutive glycine residues, at least 2, 3, 4, 5, 6, 7, 8, 9, or 10 residues in length; (2) consecutive alanine residues, at least 2, 3, 4, 5, 6, 7, 8, 9, or 10 residues in length; (3) two arginine residues (RR); (4) alanine, alanine, tyrosine (AAY); (5) a consensus sequence at least 2, 3, 4, 5, 6, 7, 8, 9, or 10 amino acid residues in length that is processed efficiently by a mammalian proteasome; and (6) one or more native sequences flanking the antigen derived from the cognate protein of origin and that is at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, or 2-20 amino acid residues in length.
260. The method of claim 257, wherein the linker links two MHC class II epitope-encoding nucleic acid sequences or an MHC class II sequence to an MHC class I epitope-encoding nucleic acid sequence.
261. The method of claim 260, wherein the linker comprises the sequence GPGPG.
262. The method of any one of the above method claims 205-223, or 225-261, wherein the antigen-encoding nucleic acid sequences is linked, operably or directly, to a separate or contiguous sequence that enhances the expression, stability, cell trafficking, processing and presentation, and/or immunogenicity of the antigen-encoding nucleic acid sequence.
263. The method of claim 262, wherein the separate or contiguous sequence comprises at least one of: a ubiquitin sequence, a ubiquitin sequence modified to increase proteasome targeting (e.g., the ubiquitin sequence contains a Gly to Ala substitution at position 76), an immunoglobulin signal sequence (e.g., IgK), a major histocompatibility class I sequence, lysosomal-associated membrane protein (LAMP)-1, human dendritic cell lysosomal-associated membrane protein, and a major histocompatibility class II sequence; optionally wherein the ubiquitin sequence modified to increase proteasome targeting is A76.
264. The method of any one of the above method claims 205-223, or 225-263, wherein the at least one antigen-encoding nucleic acid sequence comprises at least 2-10, 2, 3, 4, 5, 6, 7, 8, 9, or 10 antigen-encoding nucleic acid sequences, optionally wherein each antigen-encoding nucleic acid sequence encodes a distinct antigen-encoding nucleic acid sequence.
265. The method of any one of the above method claims 205-223, or 225-263, wherein the at least one antigen-encoding nucleic acid sequence comprises at least 11-20, 15-20, 11-

- 100, 11-200, 11-300, 11-400, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 or up to 400 antigen-encoding nucleic acid sequences, optionally wherein each antigen-encoding nucleic acid sequence encodes a distinct antigen-encoding nucleic acid sequence.
266. The composition any one of claims 205-223, or 225-263, wherein the at least one antigen-encoding nucleic acid sequence comprises at least 11-20, 15-20, 11-100, 11-200, 11-300, 11-400, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 or up to 400 antigen-encoding nucleic acid sequences.
267. The method of any one of the above method claims 205-223, or 225-263, wherein the at least one antigen-encoding nucleic acid sequence comprises at least 2-400 antigen-encoding nucleic acid sequences and wherein at least two of the antigen-encoding nucleic acid sequences encode epitope sequences or portions thereof that are presented by MHC class I on a cell surface.
268. The composition of 224, wherein at least two of the MHC class I epitopes are presented by MHC class I on the infected cell surface.
269. The method of any one of the above method claims 205-223, or 225-268, wherein the epitope-encoding nucleic acid sequences comprises at least one MHC class I epitope-encoding nucleic acid sequence, and wherein each antigen-encoding nucleic acid sequence encodes a polypeptide sequence between 8 and 35 amino acids in length, optionally 9-17, 9-25, 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 or 35 amino acids in length.
270. The method of any one of the above method claims 205-223, or 225-269, wherein the at least one MHC class II epitope-encoding nucleic acid sequence is present.
271. The method of any one of the above method claims 221-223, or 225-269, wherein the at least one MHC class II epitope-encoding nucleic acid sequence is present and comprises at least one MHC class II epitope-encoding nucleic acid sequence that comprises at least one alteration that makes the encoded epitope sequence distinct from the corresponding peptide sequence encoded by a wild-type nucleic acid sequence.
272. The method of any one of the above method claims 205-223, or 225-271, wherein the epitope-encoding nucleic acid sequence comprises an MHC class II epitope-encoding nucleic acid sequence and wherein each antigen-encoding nucleic acid sequence encodes a polypeptide sequence that is 12-20, 12, 13, 14, 15, 16, 17, 18, 19, 20, or 20-40 amino acids in length.

273. The method of any one of the above method claims 205-223, or 225-272, wherein the epitope-encoding nucleic acid sequences comprises an MHC class II epitope-encoding nucleic acid sequence, wherein the at least one MHC class II epitope-encoding nucleic acid sequence is present, and wherein the at least one MHC class II epitope-encoding nucleic acid sequence comprises at least one universal MHC class II epitope-encoding nucleic acid sequence, optionally wherein the at least one universal sequence comprises at least one of Tetanus toxoid and PADRE.
274. The method of any one of the above method claims 205, 207-223, or 225-273, wherein the at least one promoter nucleotide sequence or the second promoter nucleotide sequence is inducible.
275. The method of any one of the above method claims 205, 207-223, or 225-273, wherein the at least one promoter nucleotide sequence or the second promoter nucleotide sequence is non-inducible.
276. The method of any one of the above method claims 205, 207-223, or 225-275, wherein the at least one poly(A) sequence comprises a poly(A) sequence native to the alphavirus.
277. The method of any one of the above method claims 205, 207-223, or 225-275, wherein the at least one poly(A) sequence comprises a poly(A) sequence exogenous to the alphavirus.
278. The method of any one of the above method claims 205, 207-223, or 225-277, wherein the at least one poly(A) sequence is operably linked to at least one of the at least one nucleic acid sequences.
279. The method of any one claims 205, 207-223, or 225-278, wherein the at least one poly(A) sequence is at least 20 , at least 30, at least 40, at least 50, at least 60, at least 70, at least 80, or at least 90 consecutive A nucleotides.
280. The method of any one of the above method claims 205, 207-223, or 225-278, wherein the at least one poly(A) sequence is at least 100 consecutive A nucleotides.
281. The method of any one of the above method claims 205-223, or 225-280, wherein the epitope-encoding nucleic acid sequence comprises a MHC class I epitope-encoding nucleic acid sequence, and wherein the MHC class I epitope-encoding nucleic acid sequence is selected by performing the steps of:

- (a) obtaining at least one of exome, transcriptome, or whole genome infectious disease organism nucleotide sequencing data from the infectious disease organism, wherein the infectious disease organism nucleotide sequencing data is used to obtain data representing peptide sequences of each of a set of epitopes;
 - (b) inputting the peptide sequence of each epitope into a presentation model to generate a set of numerical likelihoods that each of the epitopes is presented by one or more of the MHC alleles on the infected cell surface, the set of numerical likelihoods having been identified at least based on received mass spectrometry data; and
 - (c) selecting a subset of the set of epitopes based on the set of numerical likelihoods to generate a set of selected epitopes which are used to generate the MHC class I epitope-encoding nucleic acid sequence.
282. The method of claim 224, wherein each of the MHC class I epitope-encoding nucleic acid sequences is selected by performing the steps of:
- (a) obtaining at least one of exome, transcriptome, or whole genome infectious disease organism nucleotide sequencing data from the infectious disease organism, wherein the infectious disease organism nucleotide sequencing data is used to obtain data representing peptide sequences of each of a set of epitopes;
 - (b) inputting the peptide sequence of each epitope into a presentation model to generate a set of numerical likelihoods that each of the epitopes is presented by one or more of the MHC alleles on the infected cell surface, the set of numerical likelihoods having been identified at least based on received mass spectrometry data; and
 - (c) selecting a subset of the set of epitopes based on the set of numerical likelihoods to generate a set of selected epitopes which are used to generate the at least 20 MHC class I epitope-encoding nucleic acid sequences.
283. The method of claim 281, wherein a number of the set of selected epitopes is 2-20.
284. The method of claim 281-283, wherein the presentation model represents dependence between:
- (a) presence of a pair of a particular one of the MHC alleles and a particular amino acid at a particular position of a peptide sequence; and

- (b) likelihood of presentation on the infected cell surface, by the particular one of the MHC alleles of the pair, of such a peptide sequence comprising the particular amino acid at the particular position.
285. The method of claim 281-284, wherein selecting the set of selected epitopes comprises selecting epitopes that have an increased likelihood of being presented on the infected cell surface relative to unselected epitopes based on the presentation model.
286. The method of claim 281-285, wherein selecting the set of selected epitopes comprises selecting epitopes that have an increased likelihood of being capable of inducing a infectious disease organism-specific immune response in the subject relative to unselected epitopes based on the presentation model.
287. The method of claim 281-286, wherein selecting the set of selected epitopes comprises selecting epitopes that have an increased likelihood of being capable of being presented to naïve T cells by professional antigen presenting cells (APCs) relative to unselected epitopes based on the presentation model, optionally wherein the APC is a dendritic cell (DC).
288. The method of claim 281-287, wherein selecting the set of selected epitopes comprises selecting epitopes that have a decreased likelihood of being subject to inhibition via central or peripheral tolerance relative to unselected epitopes based on the presentation model.
289. The method of claim 281-288, wherein selecting the set of selected epitopes comprises selecting epitopes that have a decreased likelihood of being capable of inducing an autoimmune response to normal tissue in the subject relative to unselected epitopes based on the presentation model.
290. The method of claim 281-289, wherein exome or transcriptome nucleotide sequencing data is obtained by performing sequencing on the infected tissue.
291. The method of claim 290, wherein the sequencing is next generation sequencing (NGS) or any massively parallel sequencing approach.
292. The method of any one of the above method claims 188-291, wherein the ChAdV vector comprises:
- (a) a ChAdV backbone, wherein the ChAdV backbone comprises:
- (i) at least one promoter nucleotide sequence, and

- (ii) at least one polyadenylation (poly(A)) sequence; and
- (b) a cassette, wherein the cassette comprises:
 - (i) at least one antigen-encoding nucleic acid sequence comprising:
 - a. a nucleic acid sequence encoding an infectious disease organism peptide selected from the group consisting of: a pathogen-derived peptide, a virus-derived peptide, a bacteria-derived peptide, a fungus-derived peptide, and a parasite-derived peptide,
 - b. optionally a 5' linker sequence, and
 - c. optionally a 3' linker sequence; and

wherein the cassette is operably linked to the at least one promoter nucleotide sequence and the at least one poly(A) sequence.

293. The method of any one of the above method claims 188-291, wherein the ChAdV vector comprises:

- (a) a ChAdV backbone, wherein the ChAdV backbone comprises:
 - (i) a modified ChAdV68 sequence comprising at least nucleotides 2 to 36,518 of the sequence set forth in SEQ ID NO:1, wherein the nucleotides 2 to 36,518 lack: (1) nucleotides 577 to 3403 of the sequence shown in SEQ ID NO:1 corresponding to an E1 deletion; (2) nucleotides 27,125 to 31,825 of the sequence shown in SEQ ID NO:1 corresponding to an E3 deletion; and optionally (3) nucleotides 34,916 to 35,642 of the sequence shown in SEQ ID NO:1 corresponding to a partial E4 deletion;
 - (ii) a CMV promoter nucleotide sequence; and
 - (iii) an SV40 polyadenylation (poly(A)) sequence; and
- (b) a cassette, wherein the cassette comprises:
 - (i) at least one antigen-encoding nucleic acid sequence comprising:
 - a. a nucleic acid sequence encoding an infectious disease organism peptide selected from the group consisting of: a pathogen-derived

peptide, a virus-derived peptide, a bacteria-derived peptide, a fungus-derived peptide, and a parasite-derived peptide,

b. optionally a 5' linker sequence, and

c. optionally a 3' linker sequence; and

wherein the cassette is inserted within the E1 deletion and the cassette is operably linked to the CMV promoter nucleotide sequence and the SV40 poly(A) sequence.

294. The method of any of claims 188-293, wherein the nucleic acid sequence encoding the peptide expressed in the infectious disease organism comprises an epitope-encoding nucleic acid sequence, optionally wherein the nucleic acid sequence encoding the peptide expressed in the infectious disease organism comprises two or more distinct epitope-encoding nucleic acid sequences.
295. The method of any of claims 188-293, wherein the nucleic acid sequence encoding the peptide expressed in the infectious disease organism comprises between 1-10, between 1-20, between 1-30, between 1-40, between 1-50, between 1-100, between 1-200, between 1-300, between 1-400, or between 1-500 distinct epitope-encoding nucleic acid sequences distinct epitope-encoding nucleic acid sequences, optionally wherein the nucleic acid sequence encoding the peptide expressed in the infectious disease organism comprises between 2-10, between 2-20, between 2-30, between 2-40, between 2-50, between 2-100, between 2-200, between 2-300, between 2-400, or between 2-500 distinct epitope-encoding nucleic acid sequences distinct epitope-encoding nucleic acid sequences.
296. The method of any of claims 188-293, wherein the nucleic acid sequence encoding the peptide expressed in the infectious disease organism comprises a nucleic acid sequence encoding a peptide selected from: an epitope, a full-length protein, a protein subunit, a protein domain, of the protein expressed in the infectious disease organism, optionally wherein the infectious disease organism comprises two or more distinct nucleic acid sequences encoding a peptide selected from: an epitope, a full-length protein, a protein subunit, a protein domain, and combinations thereof of the protein expressed in the infectious disease organism.

297. The method of any of claims 188-293, wherein encoded peptide or peptides is capable of stimulating an immune response when expressed in a subject, optionally wherein the immune response is a T cell response and/or a B cell response.
298. The method of any one of the above method claims, wherein the wherein the infectious disease organism is selected from the group consisting of: Severe acute respiratory syndrome-related coronavirus (SARS), severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), Ebola, HIV, Hepatitis B virus (HBV), influenza, Hepatitis C virus (HCV), Human papillomavirus (HPV), Cytomegalovirus (CMV), Chikungunya virus, Respiratory syncytial virus (RSV), Dengue virus, a orthomyxoviridae family virus, and tuberculosis.
299. The method of any one of the above method claims, wherein the epitope-encoding nucleic acid sequence encodes an epitope known or suspected to be presented by MHC class I on a surface of a cell, optionally wherein the surface of the cell is a infected cell surface or an infected cell surface, and optionally wherein the cell is the subject's cell.
300. The method of claim 299, wherein the cell is an infected cell selected from the group consisting of: a pathogen infected cell, a virally infected cell, a bacterially infected cell, an fungally infected cell, and a parasitically infected cell.
301. The method of claim 300, wherein the virally infected cell is selected from the group consisting of: an HIV infected cell, an HPV infected cell, a SARS infected cell, a SARS-CoV-2 infected cell, an Ebola infected cell, a HBV infected cell, an influenza infected cell, a HCV infected cell, a CMV infected cell, a Chikungunya virus infected cell, a RSV infected cell, a Dengue virus infected cell, a orthomyxoviridae family virus infected cell, and a tuberculosis infected cell.
302. The method of any one of the above method claims 292, or 294-301, wherein an ordered sequence of each element of the cassette in the composition for delivery of the ChAdV-based expression system is described in the formula, from 5' to 3', comprising

$$P_a-(L5_b-N_c-L3_d)_x-(G5_e-U_f)_y-G3_g$$

wherein P comprises the at least one promoter sequence operably linked to at least one of the at least one antigen-encoding nucleic acid sequences, where a = 1,

N comprises one of the epitope-encoding nucleic acid sequences, wherein the epitope-encoding nucleic acid sequence comprises an MHC class I epitope-encoding nucleic acid sequence, where $c = 1$,

L5 comprises the 5' linker sequence, where $b = 0$ or 1 ,

L3 comprises the 3' linker sequence, where $d = 0$ or 1 ,

G5 comprises one of the at least one nucleic acid sequences encoding a GPGPG amino acid linker, where $e = 0$ or 1 ,

G3 comprises one of the at least one nucleic acid sequences encoding a GPGPG amino acid linker, where $g = 0$ or 1 ,

U comprises one of the at least one MHC class II epitope-encoding nucleic acid sequence, where $f = 1$,

$X = 1$ to 400 , where for each X the corresponding N_c is an MHC class I epitope-encoding nucleic acid sequence, and

$Y = 0, 1$, or 2 , where for each Y the corresponding U_f is an MHC class II epitope-encoding nucleic acid sequence.

303. The method of claim 302, wherein for each X the corresponding N_c is a distinct MHC class I epitope-encoding nucleic acid sequence.
304. The method of claim 302 or 303, wherein for each Y the corresponding U_f is a distinct MHC class II epitope-encoding nucleic acid sequence.
305. The method of any one of the above method claims 302-304, wherein

$b = 1, d = 1, e = 1, g = 1, h = 1, X = 10, Y = 2$,

P is a CMV promoter sequence,

each N encodes a MHC class I epitope 7-15 amino acids in length,

L5 is a native 5' linker sequence that encodes a native N-terminal amino acid sequence of the MHC I epitope, and wherein the 5' linker sequence encodes a peptide that is at least 3 amino acids in length,

L3 is a native 3' linker sequence that encodes a native C-terminal amino acid sequence of the MHC I epitope, and wherein the 3' linker sequence encodes a peptide that is at least 3 amino acids in length,

U is each of a PADRE class II sequence and a Tetanus toxoid MHC class II sequence, the ChAdV vector comprises a modified ChAdV68 sequence comprising at least nucleotides 2 to 36,518 of the sequence set forth in SEQ ID NO:1, wherein the nucleotides 2 to 36,518 lack: (1) nucleotides 577 to 3403 of the sequence shown in SEQ ID NO:1 corresponding to an E1 deletion; (2) nucleotides 27,125 to 31,825 of the sequence shown in SEQ ID NO:1 corresponding to an E3 deletion; and optionally (3) nucleotides 34,916 to 35,642 of the sequence shown in SEQ ID NO:1 corresponding to a partial E4 deletion, and the antigen cassette is inserted within the E1 deletion, and each of the MHC class I antigen-encoding nucleic acid sequences encodes a polypeptide that is 25 amino acids in length.

306. The method of any one of the above method claims 292, or 294-304 wherein the cassette is integrated between the at least one promoter nucleotide sequence and the at least one poly(A) sequence.
307. The method of any one of the above method claims 292, 294-304, or 306 wherein the at least one promoter nucleotide sequence is operably linked to the cassette.
308. The method of any one of the above method claims 292, 294-304, or 306-307, wherein the ChAdV backbone comprises a ChAdV68 vector backbone.
309. The method of claim 308, wherein the ChAdV68 vector backbone comprises the sequence set forth in SEQ ID NO:1.
310. The method of claim 308, wherein the ChAdV68 vector backbone comprises a functional deletion in at least one gene selected from the group consisting of an adenovirus E1A, E1B, E2A, E2B, E3, L1, L2, L3, L4, and L5 gene with reference to a ChAdV68 genome or with reference to the sequence shown in SEQ ID NO:1, optionally wherein the adenoviral backbone or modified ChAdV68 sequence is fully deleted or functionally deleted in: (1) E1A and E1B; or (2) E1A, E1B, and E3 with

reference to the adenovirus genome or with reference to the sequence shown in SEQ ID NO:1, optionally wherein the E1 gene is functionally deleted through an E1 deletion of at least nucleotides 577 to 3403 with reference to the sequence shown in SEQ ID NO:1 and optionally wherein the E3 gene is functionally deleted through an E3 deletion of at least nucleotides 27,125 to 31,825 with reference to the sequence shown in SEQ ID NO:1.

311. The method of claim 308, wherein the ChAdV68 vector backbone comprises one or more deletions between base pair number 577 and 3403 or between base pair 456 and 3014, and optionally wherein the vector further comprises one or more deletions between base pair 27,125 and 31,825 or between base pair 27,816 and 31,333 of the sequence set forth in SEQ ID NO:1.
312. The method of claim 308, wherein the ChAdV68 vector backbone comprises at least nucleotides 2 to 36,518 of the sequence set forth in SEQ ID NO:1, wherein the nucleotides 2 to 36,518 lack:
- A. nucleotides 577 to 3403 of the sequence shown in SEQ ID NO:1 corresponding to an E1 deletion;
 - B. nucleotides 27,125 to 31,825 of the sequence shown in SEQ ID NO:1 corresponding to an E3 deletion;
 - C. nucleotides 34,916 to 35,642 of the sequence shown in SEQ ID NO:1 corresponding to a partial E4 deletion;
 - D. nucleotides 456 to 3014 of the sequence shown in SEQ ID NO:1;
 - E. nucleotides 27,816 to 31,333 of the sequence shown in SEQ ID NO:1;
 - F. nucleotides 3957 to 10346 of the sequence shown in SEQ ID NO:1;
 - G. nucleotides 21787 to 23370 of the sequence shown in SEQ ID NO:1;
 - H. nucleotides 33486 to 36193 of the sequence shown in SEQ ID NO:1; or combinations thereof.

313. The method of any one of claims 308-312, wherein the ChAdV68 vector backbone comprises a partially deleted E4 gene, optionally wherein the partially deleted E4 gene comprises:
- A. the E4 gene sequence shown in SEQ ID NO:1 and that lacks at least nucleotides 34,916 to 35,642 of the sequence shown in SEQ ID NO:1,
 - B. the E4 gene sequence shown in SEQ ID NO:1 and that lacks at least nucleotides 34,916 to 34,942, nucleotides 34,952 to 35,305 of the sequence shown in SEQ ID NO:1, nucleotides 35,302 to 35,642 of the sequence shown in SEQ ID NO:1, and wherein the vector comprises at least nucleotides 2 to 36,518 of the sequence shown in SEQ ID NO:1,
 - C. the E4 gene sequence shown in SEQ ID NO:1 and that lacks at least nucleotides 34,980 to 36,516 of the sequence shown in SEQ ID NO:1, and wherein the vector comprises at least nucleotides 2 to 36,518 of the sequence shown in SEQ ID NO:1,
 - D. the E4 gene sequence shown in SEQ ID NO:1 and that lacks at least nucleotides 34,979 to 35,642 of the sequence shown in SEQ ID NO:1, and wherein the vector comprises at least nucleotides 2 to 36,518 of the sequence shown in SEQ ID NO:1,
 - E. an E4 deletion of at least a partial deletion of E4Orf2, a fully deleted E4Orf3, and at least a partial deletion of E4Orf4,
 - F. an E4 deletion of at least a partial deletion of E4Orf2, at least a partial deletion of E4Orf3, and at least a partial deletion of E4Orf4,
 - G. an E4 deletion of at least a partial deletion of E4Orf1, a fully deleted E4Orf2, and at least a partial deletion of E4Orf3, or
 - H. an E4 deletion of at least a partial deletion of E4Orf2 and at least a partial deletion of E4Orf3.
314. The method of any one of the above method claims 292, 294-304, or 306-313, wherein the wherein the cassette is inserted in the ChAdV backbone at the E1 region, E3 region, and/or any deleted AdV region that allows incorporation of the cassette.
315. The method of any one of the above method claims 292, 294-304, or 306-314, wherein the ChAdV backbone is generated from one of a first generation, a second generation, or a helper-dependent adenoviral vector.

316. The method of any one of the above method claims claims 292, 294-304, or 306-315, wherein the at least one promoter nucleotide sequence is selected from the group consisting of: a CMV, a SV40, an EF-1, a RSV, a PGK, a HSA, a MCK, and a EBV promoter sequence.
317. The method of any one of the above method claims 292, 294-304, or 306-315, wherein the at least one promoter nucleotide sequence is a CMV promoter sequence.
318. The method of any one of the above method claims, wherein at least one of the epitope-encoding nucleic acid sequences encodes an epitope that, when expressed and translated, is capable of being presented by MHC class I on a cell of the subject.
319. The method of any one of the above method claims, wherein at least one of the epitope-encoding nucleic acid sequences encodes an epitope that, when expressed and translated, is capable of being presented by MHC class II on a cell of the subject.
320. The method of any one of the above method claims 292, 294-304, or 306-319, wherein the at least one antigen-encoding nucleic acid sequence comprises two or more antigen-encoding nucleic acid sequences.
321. The method of claim 320, wherein each antigen-encoding nucleic acid sequence is linked directly to one another.
322. The method of any one of the above method claims 292-304, or 306-321, wherein each antigen-encoding nucleic acid sequence is linked to a distinct antigen-encoding nucleic acid sequence with a nucleic acid sequence encoding a linker.
323. The method of claim 322, wherein the linker links two MHC class I epitope-encoding nucleic acid sequences or an MHC class I epitope-encoding nucleic acid sequence to an MHC class II epitope-encoding nucleic acid sequence.
324. The method of claim 323, wherein the linker is selected from the group consisting of: (1) consecutive glycine residues, at least 2, 3, 4, 5, 6, 7, 8, 9, or 10 residues in length; (2) consecutive alanine residues, at least 2, 3, 4, 5, 6, 7, 8, 9, or 10 residues in length; (3) two arginine residues (RR); (4) alanine, alanine, tyrosine (AAY); (5) a consensus sequence at least 2, 3, 4, 5, 6, 7, 8, 9, or 10 amino acid residues in length that is processed efficiently by a mammalian proteasome; and (6) one or more native sequences flanking the antigen derived from the cognate protein of origin and that is at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, or 2-20 amino acid residues in length.

325. The method of claim 322, wherein the linker links two MHC class II epitope-encoding nucleic acid sequences or an MHC class II sequence to an MHC class I epitope-encoding nucleic acid sequence.
326. The method of claim 325, wherein the linker comprises the sequence GPGPG.
327. The method of any one of the above method claims 292-304, or 306-326, wherein the antigen-encoding nucleic acid sequences is linked, operably or directly, to a separate or contiguous sequence that enhances the expression, stability, cell trafficking, processing and presentation, and/or immunogenicity of the antigen-encoding nucleic acid sequence.
328. The method of claim 327, wherein the separate or contiguous sequence comprises at least one of: a ubiquitin sequence, a ubiquitin sequence modified to increase proteasome targeting (e.g., the ubiquitin sequence contains a Gly to Ala substitution at position 76), an immunoglobulin signal sequence (e.g., IgK), a major histocompatibility class I sequence, lysosomal-associated membrane protein (LAMP)-1, human dendritic cell lysosomal-associated membrane protein, and a major histocompatibility class II sequence; optionally wherein the ubiquitin sequence modified to increase proteasome targeting is A76.
329. The method of any one of the above method claims, wherein the epitope-encoding nucleic acid sequence comprises at least one alteration that makes the encoded epitope have increased binding affinity to its corresponding MHC allele relative to the translated, corresponding wild-type nucleic acid sequence.
330. The method of any one of the above method claims, wherein the epitope-encoding nucleic acid sequence comprises at least one alteration that makes the encoded epitope have increased binding stability to its corresponding MHC allele relative to the translated, corresponding wild-type nucleic acid sequence.
331. The method of any one of the above method claims, wherein the epitope-encoding nucleic acid sequence comprises at least one alteration that makes the encoded epitope have an increased likelihood of presentation on its corresponding MHC allele relative to the translated, corresponding wild-type nucleic acid sequence.
332. The method of any one of the above method claims, wherein the at least one alteration comprises a point mutation, a frameshift mutation, a non-frameshift mutation, a deletion mutation, an insertion mutation, a splice variant, a genomic rearrangement, or a proteasome-generated spliced antigen.

333. The method of any one of the above method claims, wherein the epitope-encoding nucleic acid sequence encodes an epitope known or suspected to be expressed in the subject known or suspected to have an infection.
334. The method of claim 333, wherein the infection is selected from the group consisting of: a pathogen infection, a viral infection, a bacterial infection, an fungal infection, and a parasitic infection.
335. The method of claim 334, wherein the viral infection is selected from the group consisting of: an HIV infection, an HPV infection, a SARS infection, a SARS-CoV-2 infection, an Ebola infection, a HBV infection, an influenza infection, a HCV infection, a CMV infection, a Chikungunya virus infection, a RSV infection, a Dengue virus infection, a orthomyxoviridae family virus infection, and a tuberculosis infection.
336. The method of claim 334, wherein the bacterial infection is a tuberculosis infection.
337. The method of any one of the above method claims 292-304, or 306-336, wherein the at least one antigen-encoding nucleic acid sequence comprises at least 2-10, 2, 3, 4, 5, 6, 7, 8, 9, or 10 antigen-encoding nucleic acid sequences, optionally wherein each antigen-encoding nucleic acid sequence encodes a distinct antigen-encoding nucleic acid sequence.
338. The method of any one of the above method claims 292-304, or 306-336, wherein the at least one antigen-encoding nucleic acid sequence comprises at least 11-20, 15-20, 11-100, 11-200, 11-300, 11-400, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 or up to 400 antigen-encoding nucleic acid sequences, optionally wherein each antigen-encoding nucleic acid sequence encodes a distinct antigen-encoding nucleic acid sequence.
339. The method of any one of the above method claims 292-304, or 306-336, wherein the at least one antigen-encoding nucleic acid sequence comprises at least 11-20, 15-20, 11-100, 11-200, 11-300, 11-400, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 or up to 400 antigen-encoding nucleic acid sequences.
340. The method of any one of the above method claims 292-304, or 306-336, wherein the at least one antigen-encoding nucleic acid sequence comprises at least 2-400 antigen-encoding nucleic acid sequences and wherein at least two of the antigen-encoding nucleic acid sequences encode epitope sequences or portions thereof that are presented by MHC class I on a cell surface.
341. The composition of 340, wherein at least two of the MHC class I epitopes are presented by MHC class I on the infected cell surface.

342. The method of any one of the above method claims 292-304, or 306-341, wherein the epitope-encoding nucleic acid sequences comprises at least one MHC class I epitope-encoding nucleic acid sequence, and wherein each antigen-encoding nucleic acid sequence encodes a polypeptide sequence between 8 and 35 amino acids in length, optionally 9-17, 9-25, 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 or 35 amino acids in length.
343. The method of any one of the above method claims 292-304, or 306-342, wherein the at least one MHC class II epitope-encoding nucleic acid sequence is present.
344. The method of any one of the above method claims 292-304, or 306-342, wherein the at least one MHC class II epitope-encoding nucleic acid sequence is present and comprises at least one MHC class II epitope-encoding nucleic acid sequence that comprises at least one alteration that makes the encoded epitope sequence distinct from the corresponding peptide sequence encoded by a wild-type nucleic acid sequence.
345. The method of any one of the above method claims 292-304, or 306-344, wherein the epitope-encoding nucleic acid sequence comprises an MHC class II epitope-encoding nucleic acid sequence and wherein each antigen-encoding nucleic acid sequence encodes a polypeptide sequence that is 12-20, 12, 13, 14, 15, 16, 17, 18, 19, 20, or 20-40 amino acids in length.
346. The method of any one of the above method claims 292-304, or 306-345, wherein the epitope-encoding nucleic acid sequences comprises an MHC class II epitope-encoding nucleic acid sequence, wherein the at least one MHC class II epitope-encoding nucleic acid sequence is present, and wherein the at least one MHC class II epitope-encoding nucleic acid sequence comprises at least one universal MHC class II epitope-encoding nucleic acid sequence, optionally wherein the at least one universal sequence comprises at least one of Tetanus toxoid and PADRE.
347. The method of any one of the above method claims 292, 294-304, or 306-346, wherein the at least one promoter nucleotide sequence is inducible.
348. The method of any one of the above method claims 292, 294-304, or 306-346, wherein the at least one promoter nucleotide sequence is non-inducible.
349. The method of any one of the above method claims 292, 294-304, or 306-348, wherein the at least one poly(A) sequence comprises a Bovine Growth Hormone (BGH) SV40 polyA sequence.

350. The method of any one claims 292, 294-304, or 306-349, wherein the at least one poly(A) sequence is at least 20 , at least 30, at least 40, at least 50, at least 60, at least 70, at least 80, or at least 90 consecutive A nucleotides.
351. The method of any one of the above method claims 292, 294-304, or 306-349, wherein the at least one poly(A) sequence is at least 100 consecutive A nucleotides.
352. The method of any one of the above method claims, wherein the cassette further comprises at least one of: an intron sequence, a woodchuck hepatitis virus posttranscriptional regulatory element (WPRE) sequence, an internal ribosome entry sequence (IRES) sequence, a nucleotide sequence encoding a 2A self-cleaving peptide sequence, a nucleotide sequence encoding a Furin cleavage site, or a sequence in the 5' or 3' non-coding region known to enhance the nuclear export, stability, or translation efficiency of mRNA that is operably linked to at least one of the at least one antigen-encoding nucleic acid sequences.
353. The method of any one of the above method claims, wherein the cassette further comprises a reporter gene, including but not limited to, green fluorescent protein (GFP), a GFP variant, secreted alkaline phosphatase, luciferase, a luciferase variant, or a detectable peptide or epitope.
354. The method of claim 353, wherein the detectable peptide or epitope is selected from the group consisting of an HA tag, a Flag tag, a His-tag, or a V5 tag.
355. The method of any one of the above method claims, wherein the one or more vectors further comprises one or more nucleic acid sequences encoding at least one immune modulator.
356. The method of claim 355, wherein the immune modulator is an anti-CTLA4 antibody or an antigen-binding fragment thereof, an anti-PD-1 antibody or an antigen-binding fragment thereof, an anti-PD-L1 antibody or an antigen-binding fragment thereof, an anti-4-1BB antibody or an antigen-binding fragment thereof, or an anti-OX-40 antibody or an antigen-binding fragment thereof.
357. The method of claim 356, wherein the antibody or antigen-binding fragment thereof is a Fab fragment, a Fab' fragment, a single chain Fv (scFv), a single domain antibody (sdAb) either as single specific or multiple specificities linked together (e.g., camelid antibody domains), or full-length single-chain antibody (e.g., full-length IgG with heavy and light chains linked by a flexible linker).

358. The method of claim 356 or 357, wherein the heavy and light chain sequences of the antibody are a contiguous sequence separated by either a self-cleaving sequence such as 2A or IRES; or the heavy and light chain sequences of the antibody are linked by a flexible linker such as consecutive glycine residues.
359. The method of claim 355, wherein the immune modulator is a cytokine.
360. The method of claim 359, wherein the cytokine is at least one of IL-2, IL-7, IL-12, IL-15, or IL-21 or variants thereof of each.
361. The method of any one of the above method claims 292-304, or 306-360, wherein the epitope-encoding nucleic acid sequence comprises a MHC class I epitope-encoding nucleic acid sequence, and wherein the MHC class I epitope-encoding nucleic acid sequence is selected by performing the steps of:
- (a) obtaining at least one of exome, transcriptome, or whole genome infectious disease organism nucleotide sequencing data from the infectious disease organism, wherein the infectious disease organism nucleotide sequencing data is used to obtain data representing peptide sequences of each of a set of epitopes;
 - (b) inputting the peptide sequence of each epitope into a presentation model to generate a set of numerical likelihoods that each of the epitopes is presented by one or more of the MHC alleles on the infected cell surface, the set of numerical likelihoods having been identified at least based on received mass spectrometry data; and
 - (c) selecting a subset of the set of epitopes based on the set of numerical likelihoods to generate a set of selected epitopes which are used to generate the MHC class I epitope-encoding nucleic acid sequence.
362. The method of claim 305, wherein each of the MHC class I epitope-encoding nucleic acid sequences is selected by performing the steps of:
- (a) obtaining at least one of exome, transcriptome, or whole genome infectious disease organism nucleotide sequencing data from the infectious disease organism, wherein the infectious disease organism nucleotide sequencing data is used to obtain data representing peptide sequences of each of a set of epitopes;

- (b) inputting the peptide sequence of each epitope into a presentation model to generate a set of numerical likelihoods that each of the epitopes is presented by one or more of the MHC alleles on the infected cell surface, the set of numerical likelihoods having been identified at least based on received mass spectrometry data; and
 - (c) selecting a subset of the set of epitopes based on the set of numerical likelihoods to generate a set of selected epitopes which are used to generate the at least 20 MHC class I epitope-encoding nucleic acid sequences.
363. The method of claim 361, wherein a number of the set of selected epitopes is 2-20.
364. The method of any one of claims 361-363, wherein the presentation model represents dependence between:
- (a) presence of a pair of a particular one of the MHC alleles and a particular amino acid at a particular position of a peptide sequence; and
 - (b) likelihood of presentation on the infected cell surface, by the particular one of the MHC alleles of the pair, of such a peptide sequence comprising the particular amino acid at the particular position.
365. The method of any one of claims 361-364, wherein selecting the set of selected epitopes comprises selecting epitopes that have an increased likelihood of being presented on the infected cell surface relative to unselected epitopes based on the presentation model.
366. The method of any one of claims 361-365, wherein selecting the set of selected epitopes comprises selecting epitopes that have an increased likelihood of being capable of inducing a infectious disease organism-specific immune response in the subject relative to unselected epitopes based on the presentation model.
367. The method of any one of claims 361-366, wherein selecting the set of selected epitopes comprises selecting epitopes that have an increased likelihood of being capable of being presented to naïve T cells by professional antigen presenting cells (APCs) relative to unselected epitopes based on the presentation model, optionally wherein the APC is a dendritic cell (DC).
368. The method of any one of claims 361-367, wherein selecting the set of selected epitopes comprises selecting epitopes that have a decreased likelihood of being subject to inhibition via central or peripheral tolerance relative to unselected epitopes based on the presentation model.

369. The method of any one of claims 361-368, wherein selecting the set of selected epitopes comprises selecting epitopes that have a decreased likelihood of being capable of inducing an autoimmune response to normal tissue in the subject relative to unselected epitopes based on the presentation model.
370. The method of any one of claims 361-369, wherein exome or transcriptome nucleotide sequencing data is obtained by performing sequencing on the infected tissue.
371. The method of claim 370, wherein the sequencing is next generation sequencing (NGS) or any massively parallel sequencing approach.
372. The method of any of the above claims, wherein the cassette comprises junctional epitope sequences formed by adjacent sequences in the cassette.
373. The method of claim 372, wherein at least one or each junctional epitope sequence has an affinity of greater than 500 nM for MHC.
374. The method of claims 372 or 373, wherein each junctional epitope sequence is non-self.
375. The method of any of the above claims, wherein the cassette does not encode a non-therapeutic MHC class I or class II epitope nucleic acid sequence comprising a translated, wild-type nucleic acid sequence, wherein the non-therapeutic epitope is predicted to be displayed on an MHC allele of the subject.
376. The method of claim 375, wherein the non-therapeutic predicted MHC class I or class II epitope sequence is a junctional epitope sequence formed by adjacent sequences in the cassette.
377. The method of claims 372-376, wherein the prediction is based on presentation likelihoods generated by inputting sequences of the non-therapeutic epitopes into a presentation model.
378. The method of any one of the above method claims 372-377, wherein an order of the antigen-encoding nucleic acid sequences in the cassette is determined by a series of steps comprising:
- (a) generating a set of candidate cassette sequences corresponding to different orders of the antigen-encoding nucleic acid sequences;
 - (b) determining, for each candidate cassette sequence, a presentation score based on presentation of non-therapeutic epitopes in the candidate cassette sequence; and
 - (c) selecting a candidate cassette sequence associated with a presentation score below a predetermined threshold as the cassette sequence for a vaccine.

379. The method of any of the above claims, wherein the composition for delivery of the ChAdV-based expression system is formulated in a pharmaceutical composition comprising a pharmaceutically acceptable carrier.
380. The method of any of the above claims, wherein one or more of the epitope-encoding nucleic acid sequences are derived from an infection in or an infected cell of the subject.
381. The method of any of the above claims, wherein each of the epitope-encoding nucleic acid sequences are derived from an infection in or an infected cell of the subject.
382. The method of any of the above claims, wherein one or more of the epitope-encoding nucleic acid sequences are not derived from an infection in or an infected cell of the subject.
383. The method of any of the above claims, wherein each of the epitope-encoding nucleic acid sequences are not derived from an infection in or an infected cell of the subject.
384. A method for stimulating an immune response in a subject, the method comprising administering to the subject a composition for delivery of a self-replicating alphavirus-based expression system:

wherein the composition for delivery of the self-replicating alphavirus-based expression system comprises:

(A) the self-replicating alphavirus-based expression system, wherein the self-replicating alphavirus-based expression system comprises one or more vectors, wherein the one or more vectors comprises:

(a) an RNA alphavirus backbone, wherein the RNA alphavirus backbone comprises:

(i) at least one promoter nucleotide sequence, and

(ii) at least one polyadenylation (poly(A)) sequence; and

(b) a cassette, wherein the cassette comprises:

(i) at least one antigen-encoding nucleic acid sequence comprising:

a. a nucleic acid sequence encoding an infectious disease organism peptide selected from the group consisting of: a pathogen-derived

peptide, a virus-derived peptide, a bacteria-derived peptide, a fungus-derived peptide, and a parasite-derived peptide, wherein the infectious disease organism is selected from the group consisting of: Severe acute respiratory syndrome-related coronavirus (SARS), severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), Ebola, HIV, Hepatitis B virus (HBV), influenza, Hepatitis C virus (HCV), Human papillomavirus (HPV), Cytomegalovirus (CMV), Chikungunya virus, Respiratory syncytial virus (RSV), Dengue virus, a orthomyxoviridae family virus, and tuberculosis;

b. optionally a 5' linker sequence, and

c. optionally a 3' linker sequence;

(ii) optionally, a second promoter nucleotide sequence operably linked to the at least one antigen-encoding nucleic acid sequence; and

(iii) optionally, at least one second poly(A) sequence, wherein the second poly(A) sequence is a native poly(A) sequence or an exogenous poly(A) sequence to the alphavirus, and

(B) a lipid-nanoparticle (LNP), wherein the LNP encapsulates the self-replicating alphavirus-based expression system

385. A method for stimulating an immune response in a subject, the method comprising administering to the subject a composition for delivery of a chimpanzee adenovirus (ChAdV)-based expression system, wherein the composition for delivery of the ChAdV-based expression system comprises:

the ChAdV-based expression system, wherein the ChAdV-based expression system comprises a viral particle comprising a ChAdV vector, wherein the ChAdV vector comprises:

(a) a ChAdV backbone, wherein the ChAdV backbone comprises:

(i) at least one promoter nucleotide sequence, and

(ii) at least one polyadenylation (poly(A)) sequence; and

- (b) a cassette, wherein the cassette comprises:
 - (i) at least one antigen-encoding nucleic acid sequence comprising:
 - a. a nucleic acid sequence encoding an infectious disease organism peptide selected from the group consisting of: a pathogen-derived peptide, a virus-derived peptide, a bacteria-derived peptide, a fungus-derived peptide, and/or a parasite-derived peptide, wherein the infectious disease organism is selected from the group consisting of: Severe acute respiratory syndrome-related coronavirus (SARS), severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), Ebola, HIV, Hepatitis B virus (HBV), influenza, Hepatitis C virus (HCV), Human papillomavirus (HPV), Cytomegalovirus (CMV), Chikungunya virus, Respiratory syncytial virus (RSV), Dengue virus, a orthomyxoviridae family virus, and tuberculosis;
 - b. optionally a 5' linker sequence, and
 - c. optionally a 3' linker sequence; and

wherein the cassette is operably linked to the at least one promoter nucleotide sequence and the at least one poly(A) sequence.

- 386. The method of any of the above claims 384 or 385, wherein the self-replicating alphavirus-based expression system or the ChAdV-based expression system comprise any of the features of any one of the above claims.
- 387. The method of any of the above claims, wherein the cassette of the composition for delivery of the ChAdV-based expression system is identical to the cassette of the composition for delivery of the self-replicating alphavirus-based expression system.

VEE-MAG25mer srRNA-LNP1(30 µg)

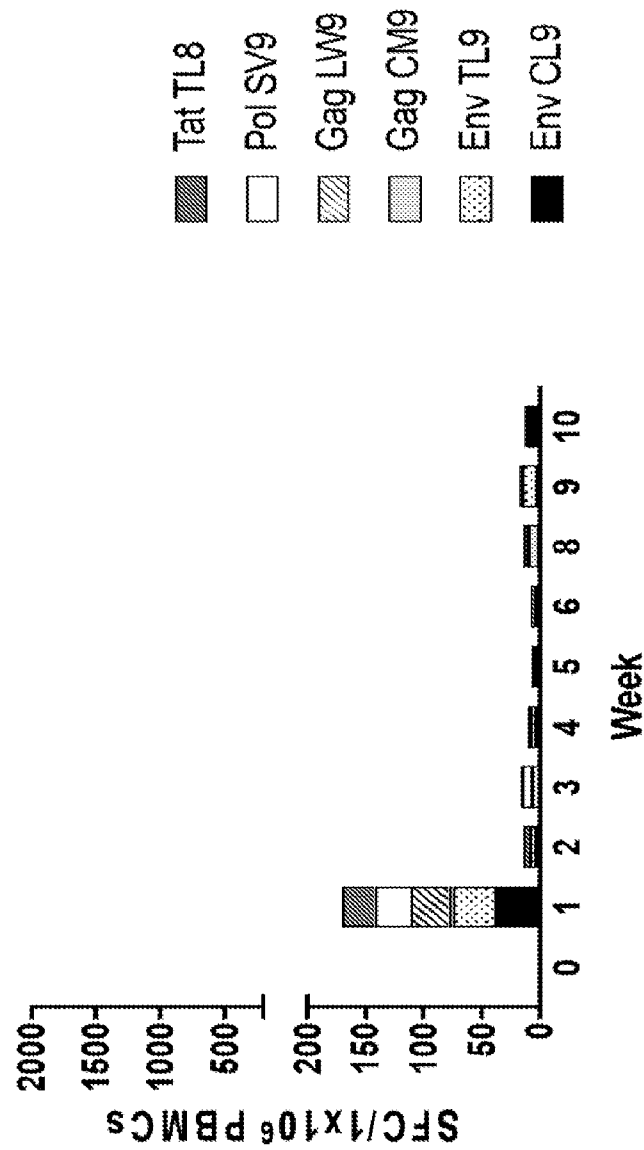


FIG 1A

VEE-MAG25mer srRNA-LNP1(100 µg)

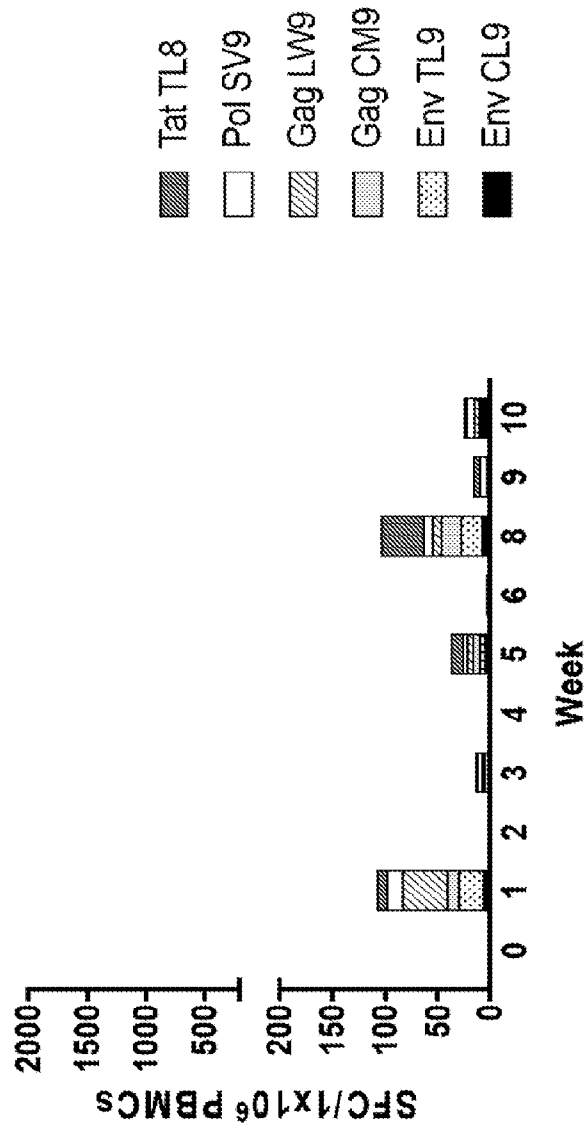


FIG 1B

VEE-MAG25mer srRNA-LNP2(100 µg)

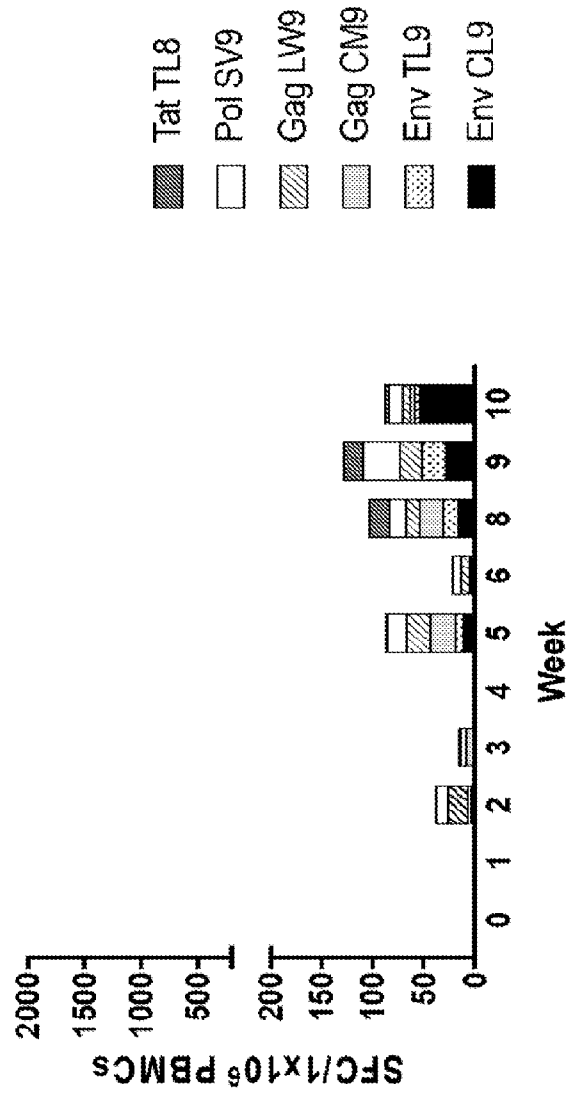


FIG 1C

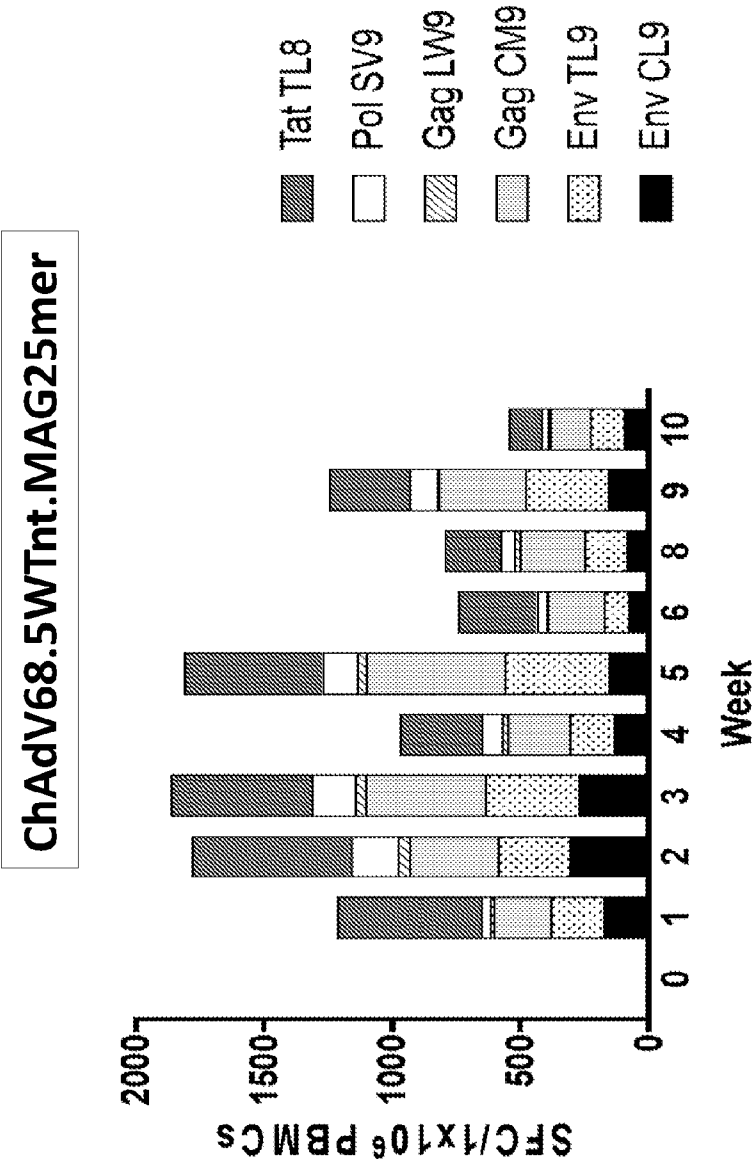


FIG 1D

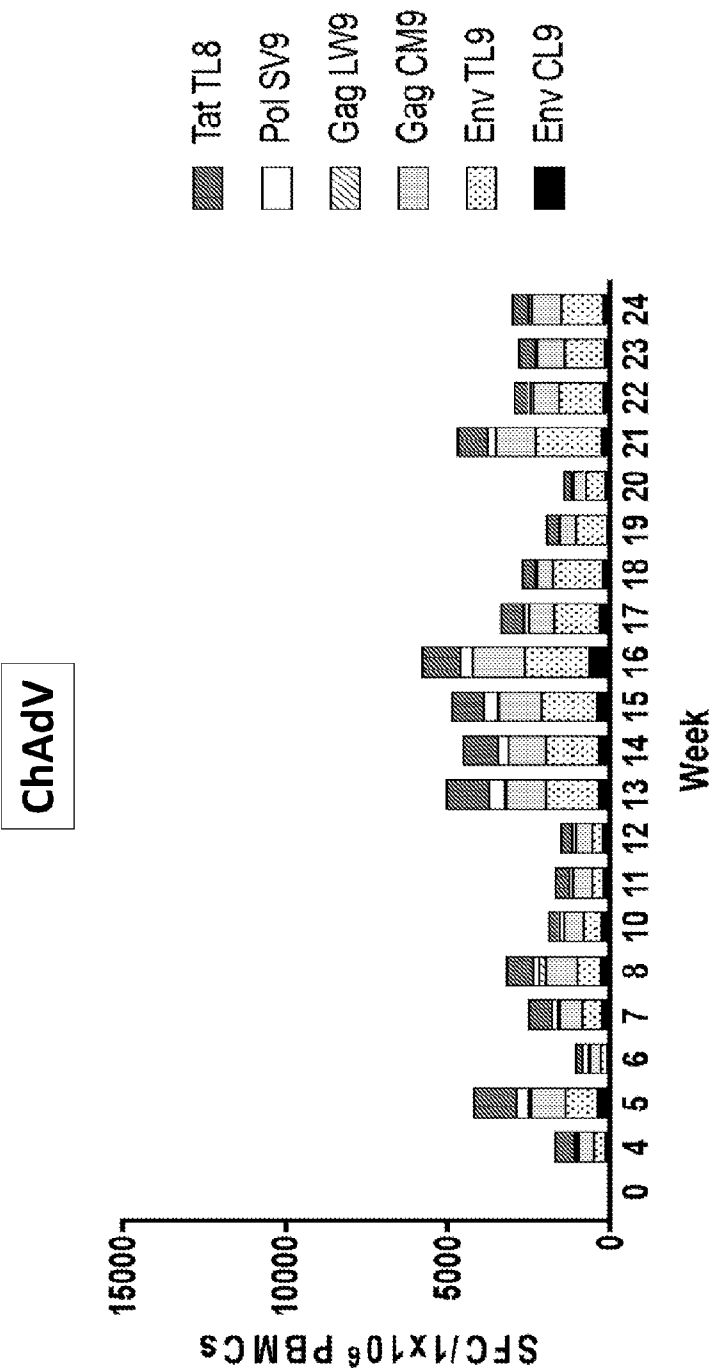


FIG 2

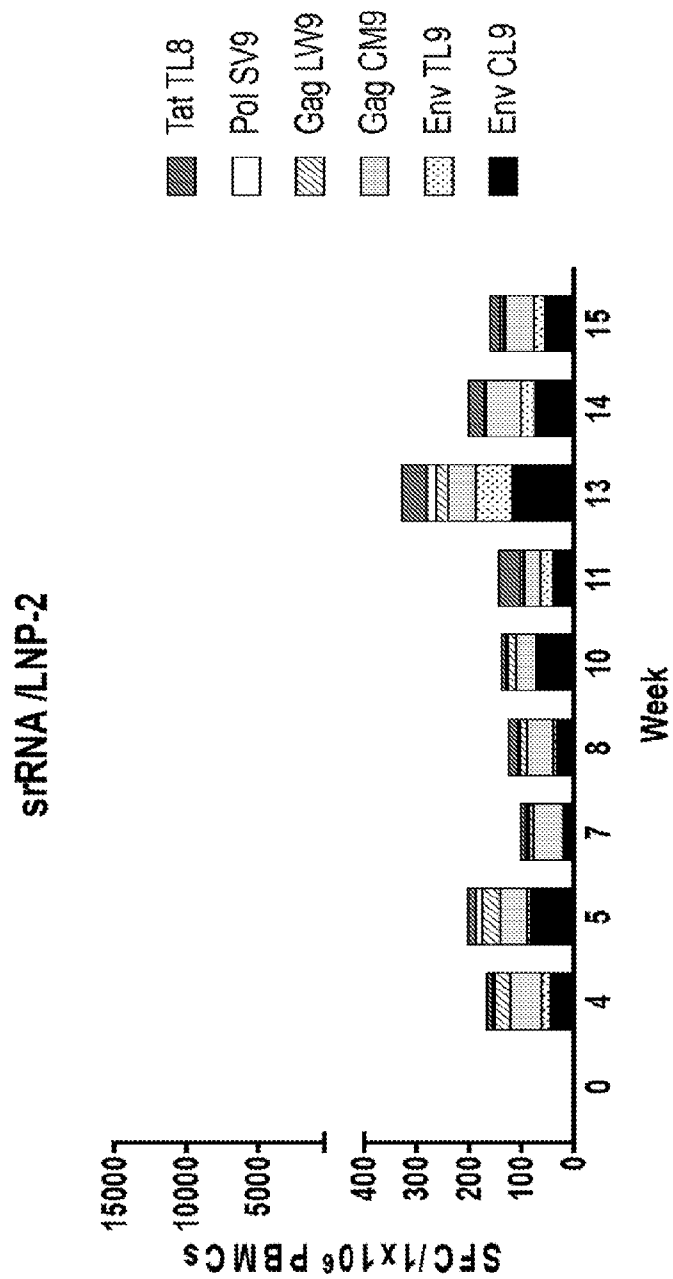


FIG 3

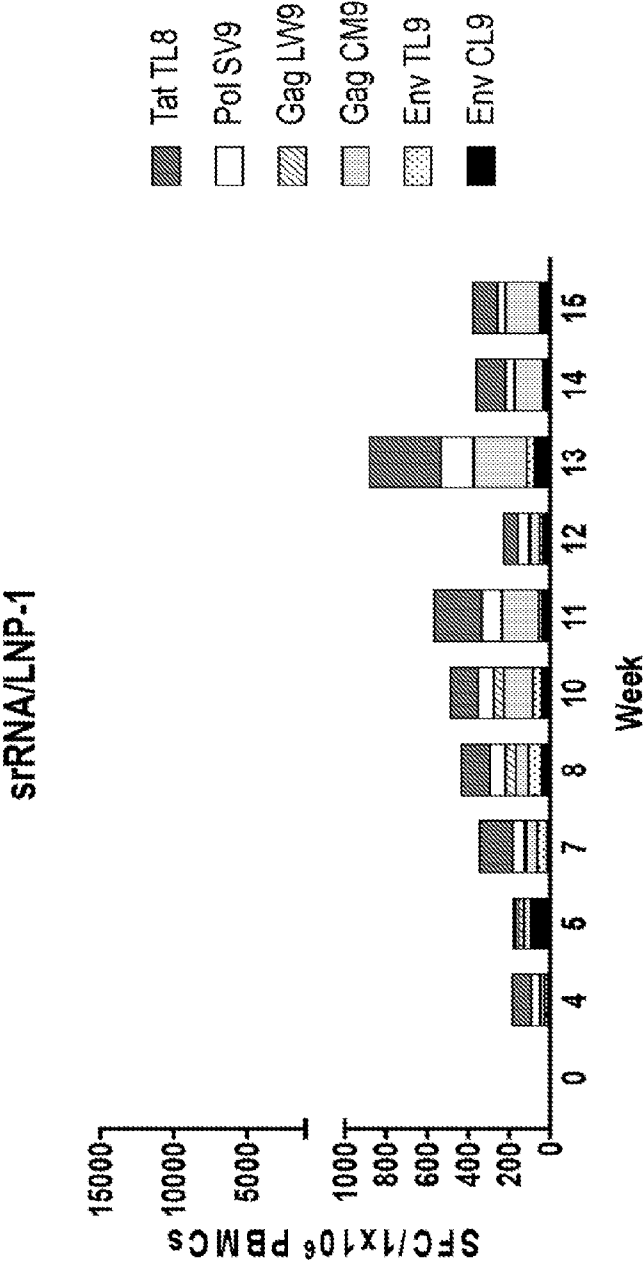


FIG 4

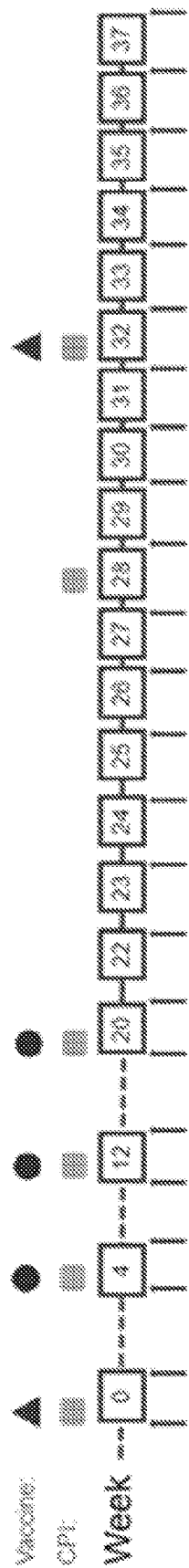


FIG 5

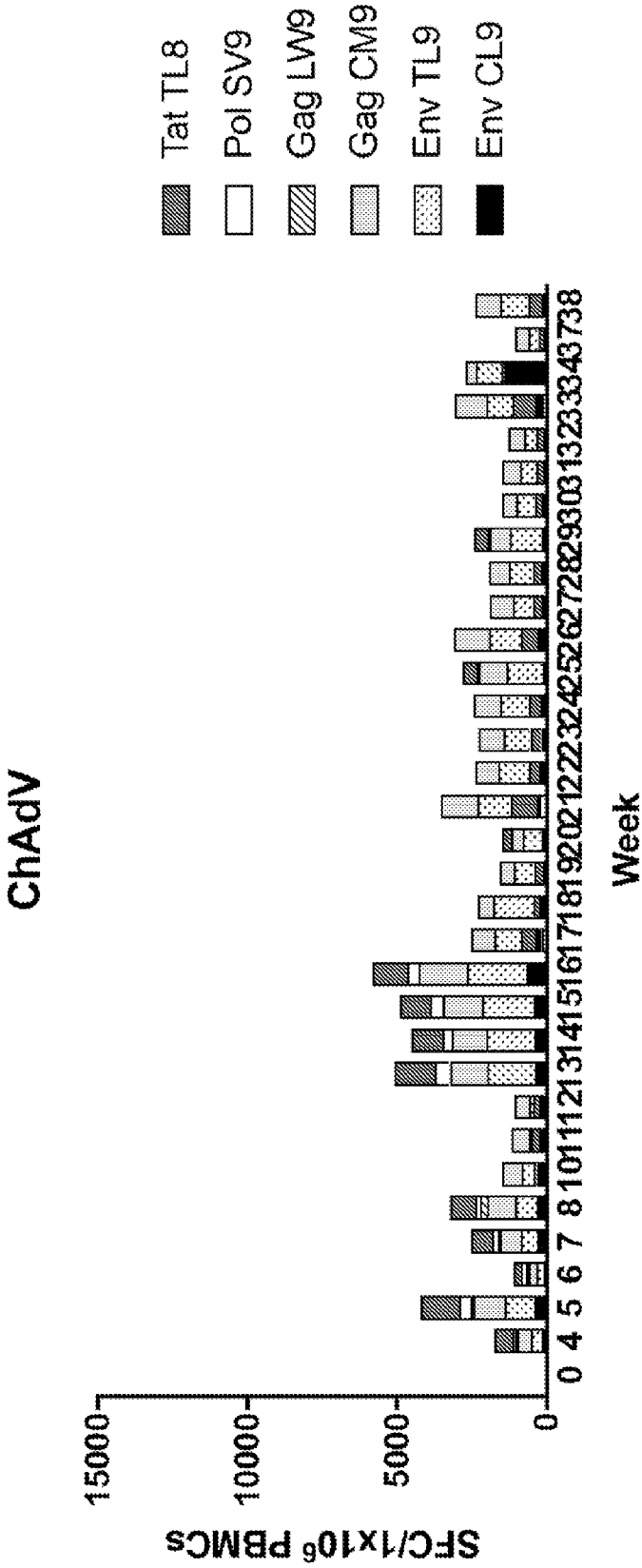


FIG 6

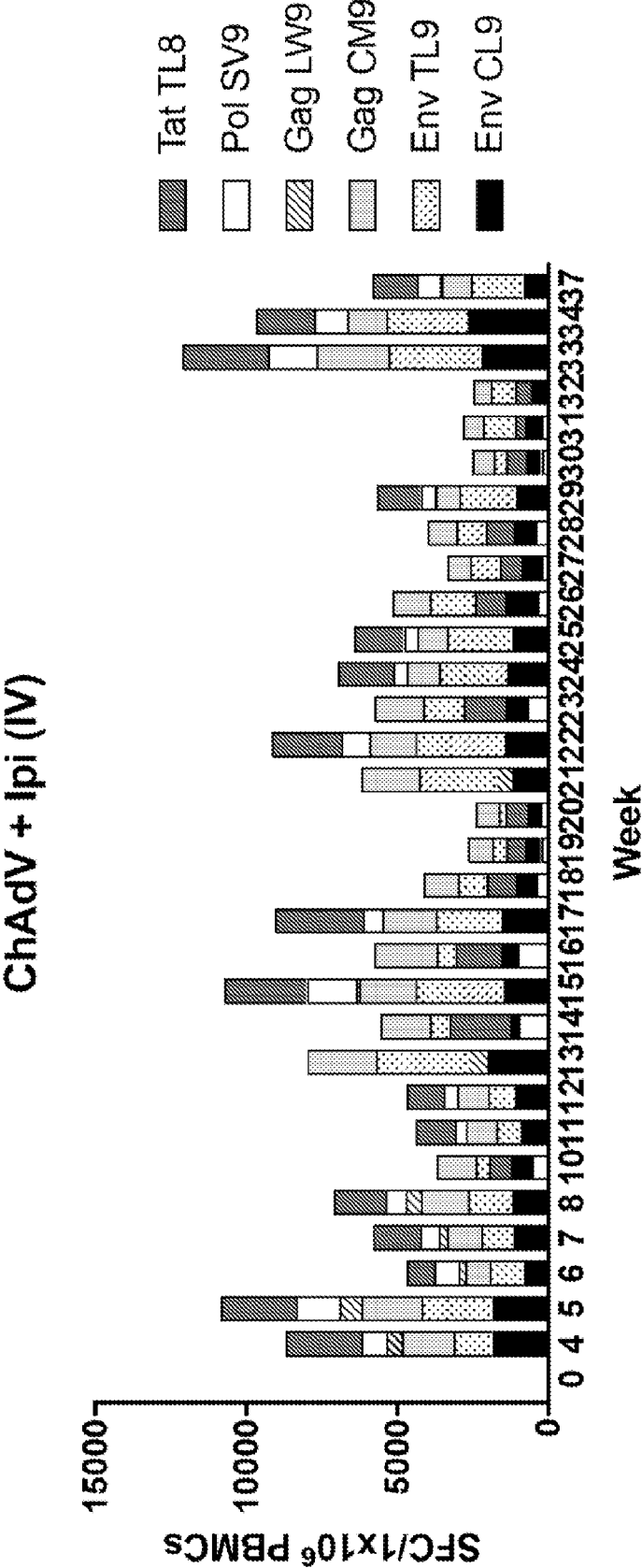


FIG 7

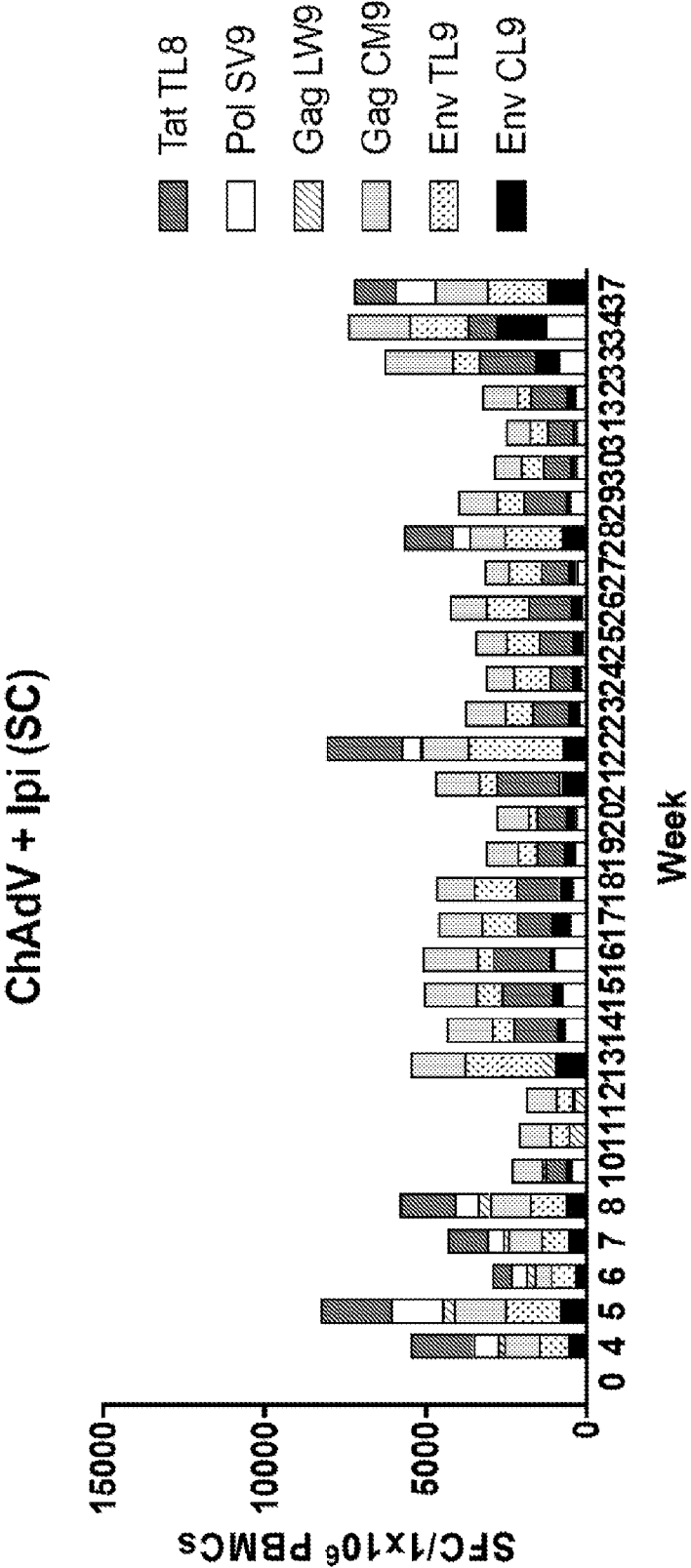


FIG 8

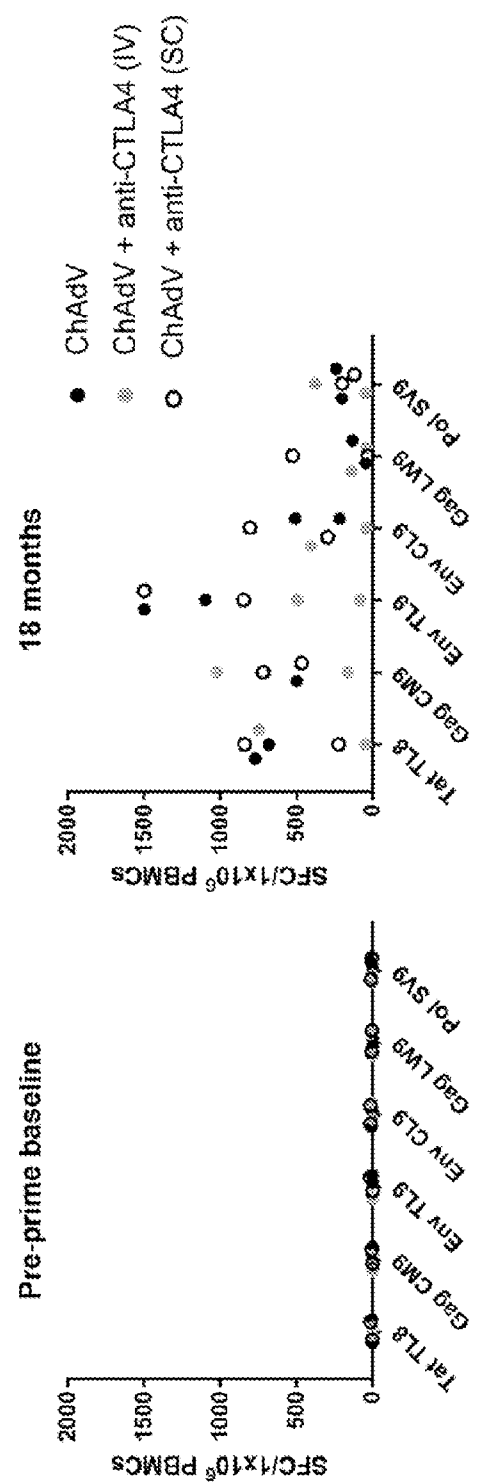


FIG 9

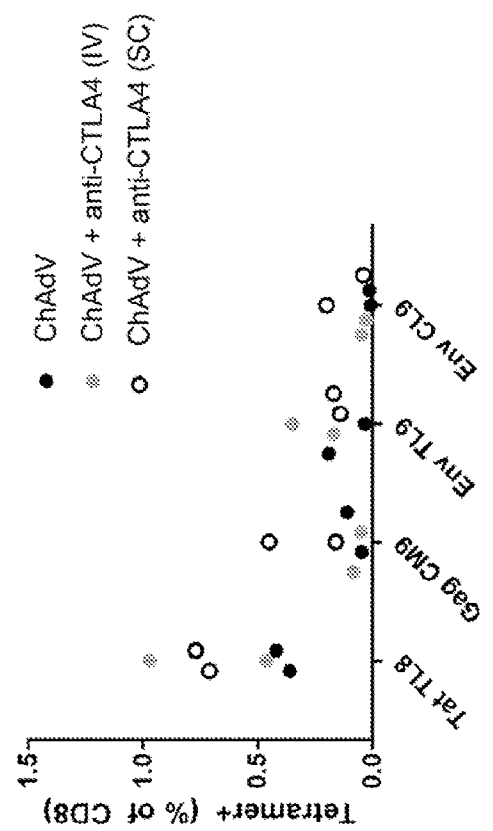


FIG 10

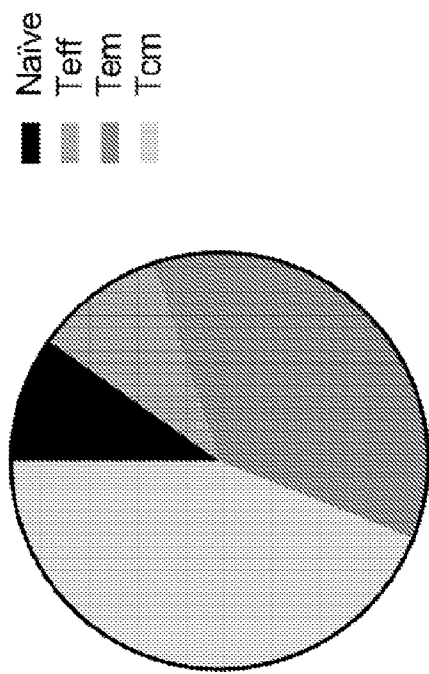


FIG 11

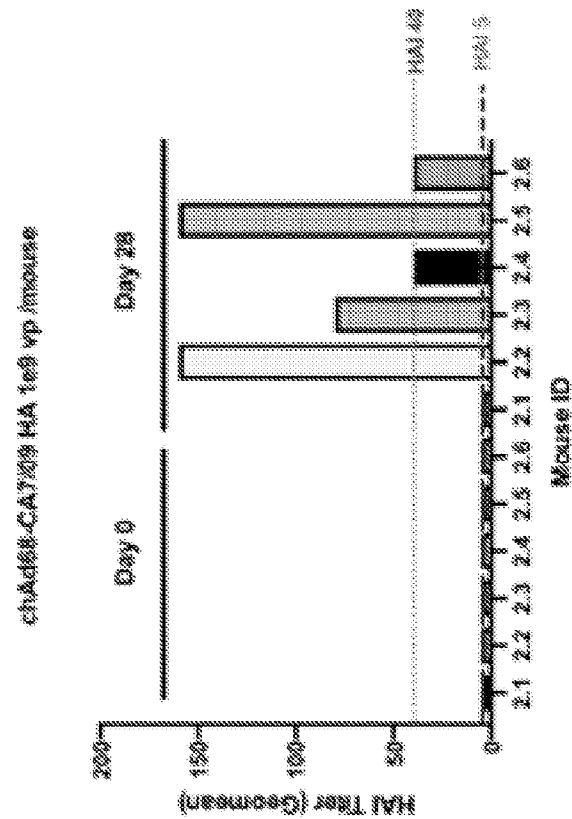


FIG 12B

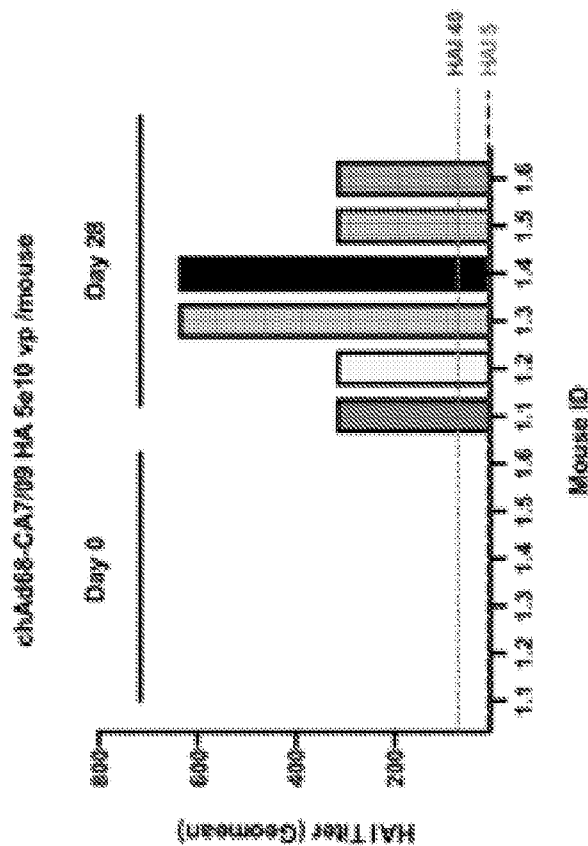


FIG 12A

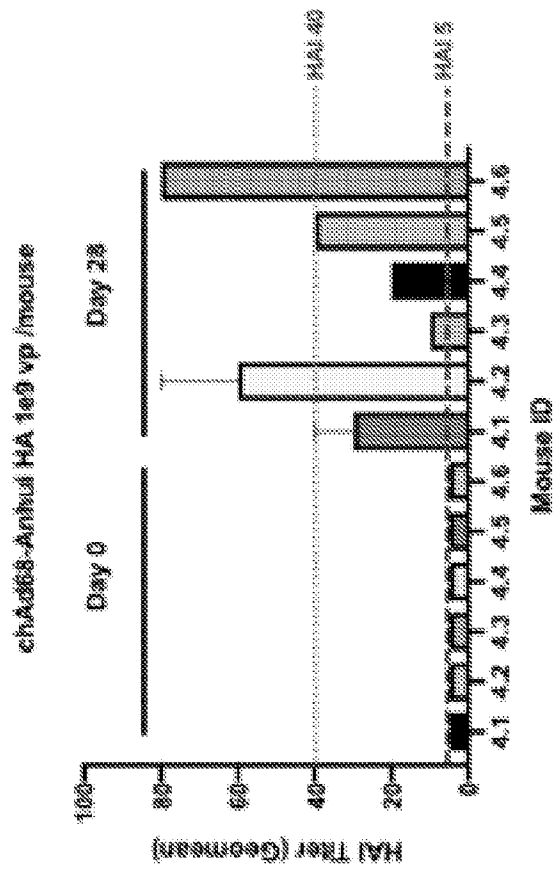


FIG 13B

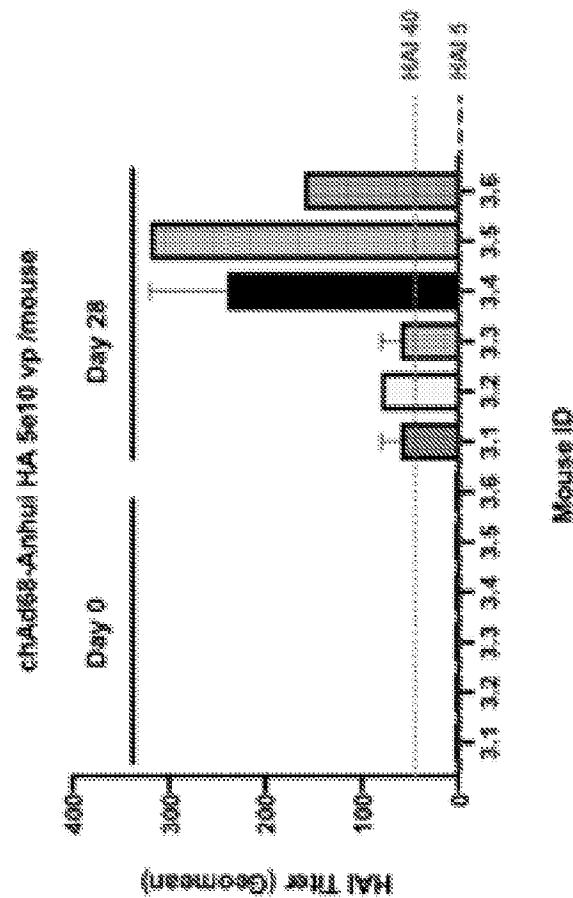


FIG 13A

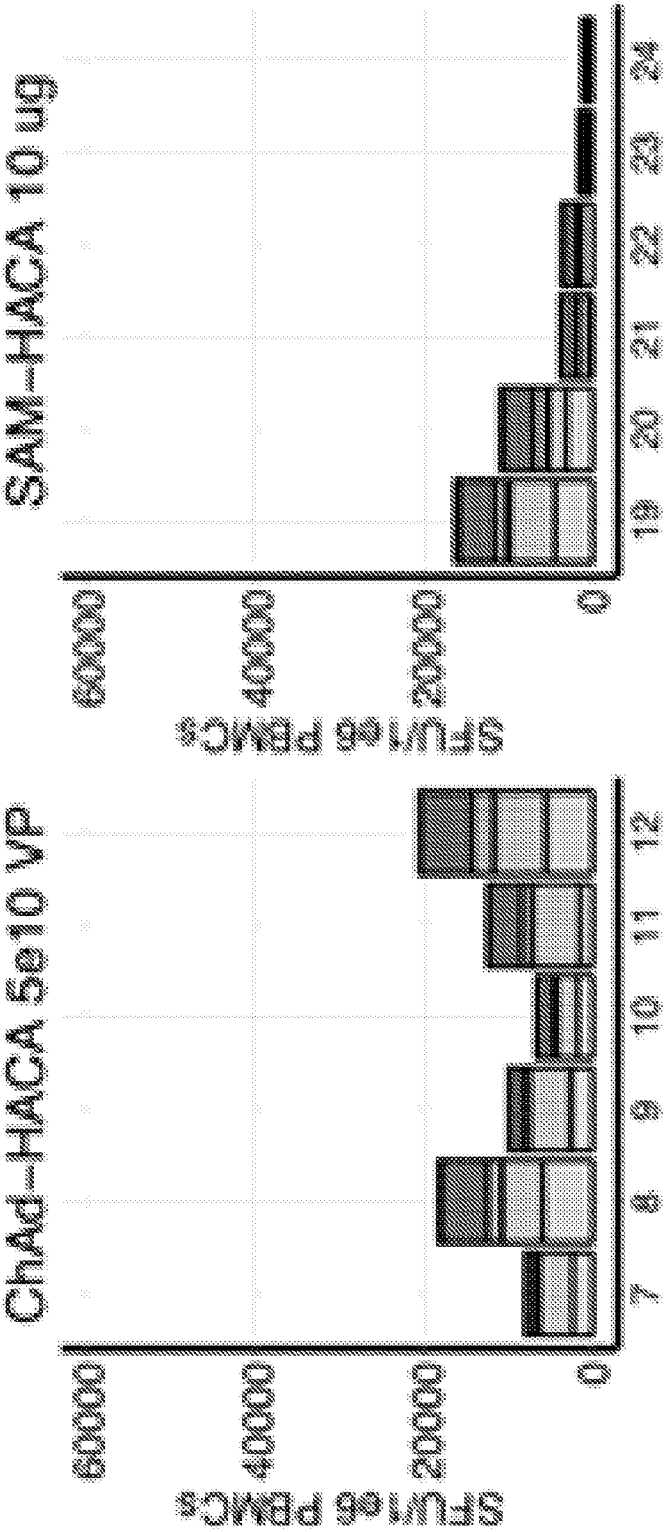


FIG 14B

FIG 14A

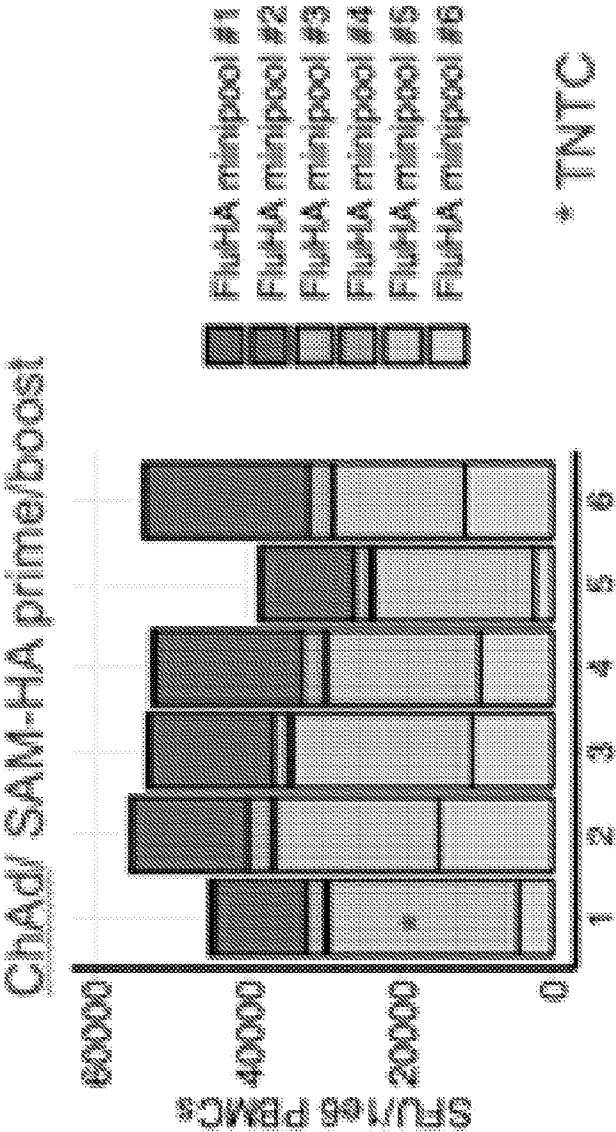


FIG 14C

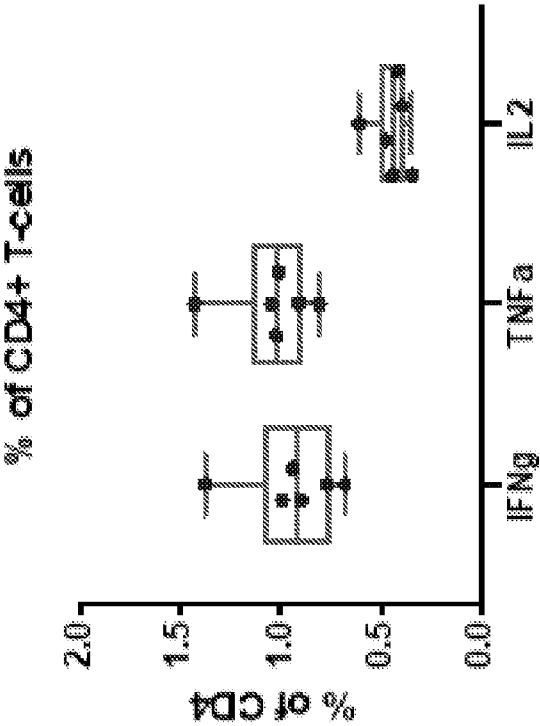


FIG 15B

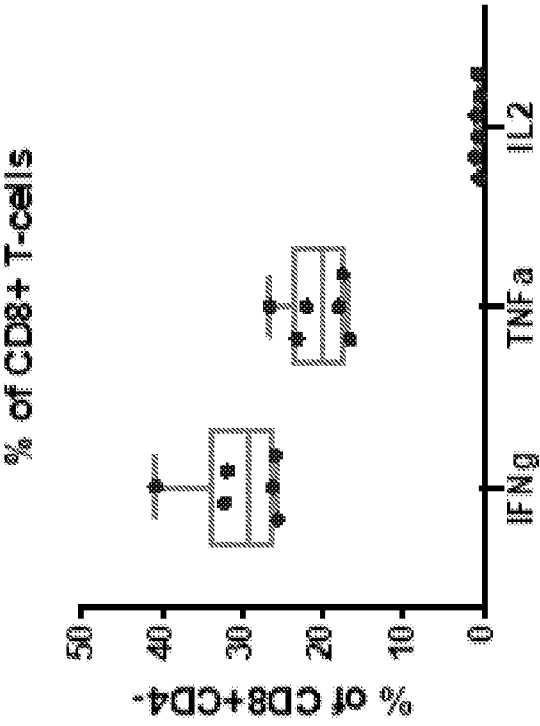


FIG 15A

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2021/025828

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. ☒ forming part of the international application as filed:
☒ in the form of an Annex C/ST.25 text file.
☐ on paper or in the form of an image file.
- b. ☐ furnished together with the international application under PCT Rule 13ter.1(a) for the purposes of international search only in the form of an Annex C/ST.25 text file.
- c. ☐ furnished subsequent to the international filing date for the purposes of international search only:
☐ in the form of an Annex C/ST.25 text file (Rule 13ter.1(a)).
☐ on paper or in the form of an image file (Rule 13ter.1(b) and Administrative Instructions, Section 713).
2. ☐ In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that forming part of the application as filed or does not go beyond the application as filed, as appropriate, were furnished.
3. Additional comments:

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2021/025828

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☒ Claims Nos.: 12-89, 95-184, 187, 193-383, 386, 387
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.
- ☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.
- ☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2021/025828

A. CLASSIFICATION OF SUBJECT MATTER

IPC(8) - A61K 35/00; A61K 35/76; A61K 38/04; A61K 39/00; A61K 39/12; A61K 48/00 (2021.01)

CPC - A61K 35/00; A61K 35/76; A61K 39/00; A61K 39/12; A61K 48/00; C12N 2770/36111; C12N 2770/36134 (2021.05)

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)

see Search History document

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

see Search History document

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

see Search History document

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 2018/208856 A1 (GRITSTONE ONCOLOGY, INC.) 15 November 2018 (15.11.2018) entire document	1-11, 90-94, 185, 186, 188-192, 384, 385
A	WO 2016/122414 A1 (AGENCY FOR SCIENCE, TECHNOLOGY AND RESEARCH et al) 04 August 2016 (04.08.2016) entire document	1-11, 90-94, 185, 186, 188-192, 384, 385
A	US 8,216,834 B2 (COLLOCA et al) 10 July 2012 (10.07.2012) entire document	1-11, 90-94, 185, 186, 188-192, 384, 385
A	US 6,376,236 B1 (DUBENSKY, JR. et al) 23 April 2002 (23.04.2002) entire document	1-11, 90-94, 185, 186, 188-192, 384, 385
A	US 6,770,283 B1 (GAROFF et al) 03 August 2004 (03.08.2004) entire document	1-11, 90-94, 185, 186, 188-192, 384, 385
A	WO 2005/071093 A2 (ISTITUTO DI RICERCHE DI BIOLOGIA MOLECOLARE P. ANGELETTI S.P.A. et al) 04 August 2005 (04.08.2005) entire document	1-11, 90-94, 185, 186, 188-192, 384, 385



Further documents are listed in the continuation of Box C.



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

"D" document cited by the applicant in the international application

"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

"I" 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

30 June 2021

Date of mailing of the international search report

28 JUL 2021

Name and mailing address of the ISA/US

Mail Stop PCT, Attn: ISA/US, Commissioner for Patents
P.O. Box 1450, Alexandria, VA 22313-1450

Facsimile No. 571-273-8300

Authorized officer

Harry Kim

Telephone No. PCT Helpdesk: 571-272-4300