



- (51) International Patent Classification:
C12N 15/86 (2006.01) A61K 48/00 (2006.01)
- (21) International Application Number:
PCT/US2022/079505
- (22) International Filing Date:
08 November 2022 (08.11.2022)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
63/277,116 08 November 2021 (08.11.2021) US
- (71) Applicant: GRITSTONE BIO, INC. [US/US]; 5959 Horton Street, Suite 300, Emeryville, California 94608 (US).
- (72) Inventors: HONG, Sue-Jean; c/o Gritstone Bio, Inc., 5959 Horton Street, Suite 300, Emeryville, California 94608 (US). FIDANZA, Mario; c/o Gritstone Bio, Inc., 5959 Horton Street, Suite 300, Emeryville, California 94608 (US). JOOSS, Karin; c/o Gritstone Bio, 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, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.
- (84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, 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, ME, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI,

(54) Title: SELF-AMPLIFYING RNA COMPOSITIONS AND METHODS OF USE THEREOF

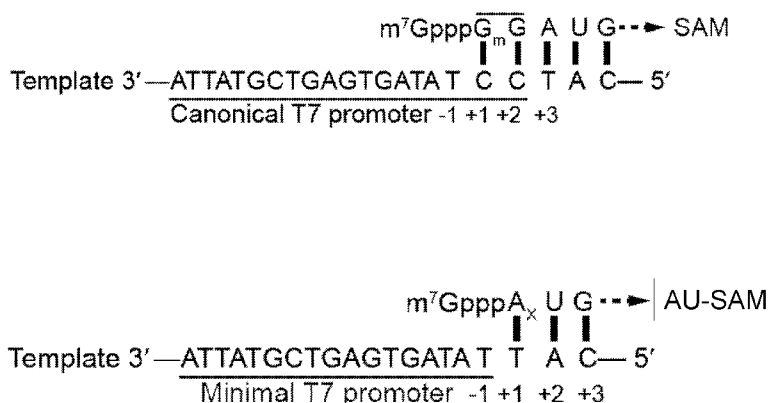


FIG. 1

(57) Abstract: The present disclosure includes, among other things, RNA polynucleotide compositions that (A) include modified nucleosides; and/or (B) have been purified using a chromatography system and/or an affinity-based separation system. Methods of production and treatment related to the same are also provided.



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:

- *without international search report and to be republished upon receipt of that report (Rule 48.2(g))*

SELF-AMPLIFYING RNA COMPOSITIONS AND METHODS OF USE THEREOF**Cross reference to related applications**

[0001] This application claims the benefit of U.S. Provisional Application No. 63/277,116 filed November 8, 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 ###, is named ### and is ### bytes in size.

Background

[0003] Messenger RNA (mRNA), encoding physiologically important proteins for therapeutic applications, has shown significant advantages over DNA-based plasmid and viral vectors for delivering genetic material. However, mRNA-based therapies can have potential drawbacks, such as off-target stimulation of the innate immune response that may interfere with therapeutic efficacy.

[0004] Accordingly, there is a need in the industry for improved mRNA therapies and production methods.

Summary

[0005] Provided herein is a composition for delivery of a self-amplifying expression system comprising a single-stranded RNA (ssRNA) vector, wherein the ssRNA vector comprises a 7-methylguanylate (m⁷G cap), a polyadenylation (polyA) tail, a self-amplifying backbone, and one or more modified nucleosides, wherein the one or more modified nucleosides comprise m5C and no other modified nucleotides, optionally other than an m⁷G cap analog

[0006] Also provided herein is a composition for delivery of a self-amplifying expression system comprising a single-stranded RNA (ssRNA) vector, wherein the ssRNA vector comprises a 7-methylguanylate (m⁷G cap), a polyadenylation (polyA) tail, a self-amplifying backbone, and one or more modified nucleosides.

[0007] In some aspects, the self-amplifying backbone comprises a polynucleotide selected from a self-replicating RNA virus. In some aspects, the self-replicating RNA virus is selected from the group comprising an alphavirus, a flavivirus, a measles virus, and a rhabdovirus.

[0008] In some aspects, the one or more modified nucleosides comprise methyl-5-cytosine (m5C), methyl-6-adenosine (m6A), ribose-methylation (2'-O-Me), s-thiouridine, (s2U), 5-methyluridine (m5U), N1-methyl pseudouridine (m1Ψ), pseudouridine (Ψ), or any combination

thereof. In some aspects, the one or more modified nucleosides comprise m⁵C. In some aspects, the one or more modified nucleosides comprise m⁵C and no other modified nucleotides, optionally other than an m⁷G cap analog.

[0009] In some aspects, the ssRNA vector is purified. In some aspects, the ssRNA vector is purified by chromatography. In some aspects, the chromatography comprises a cellulose chromatography system or an affinity-based separation system. In some aspects, the affinity-based separation system is a deoxythymidine (dT) oligonucleotide (Oligo(dT)) system.

[0010] In some aspects, greater than 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, or 100% of adenine, guanine, cytidine, and/or uridine nucleosides of the ssRNA vector are modified nucleosides. In some aspects, greater than 25% of adenosine, guanosine, cytidine, and/or uridine nucleosides of the ssRNA vector comprise modified nucleosides. In some aspects, 100% of adenine, guanine, cytidine, and/or uridine nucleosides of the ssRNA vector comprise modified nucleosides. In some aspects, the ssRNA vector comprises 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, or 100% of cytidine nucleosides comprise modified nucleosides. In some aspects, greater than 25% of cytidine nucleosides of the ssRNA vector comprise modified nucleosides. In some aspects, the ssRNA comprises 100% of cytidine nucleosides are modified nucleosides.

[0011] In some aspects, the composition comprises less than 30%, less than 25%, less than 20%, less than 15%, less than 10%, less than 5%, less than 2%, less than 1%, less than 0.5%, or less than 0.1% of a contaminant. In some aspects, the contaminant comprises salt, detergent, and/or double stranded RNA (dsRNA). In some aspects, the contaminant comprises dsRNA.

[0012] In some aspects, the ssRNA vector comprises 70% or more, 80% or more, 85% or more, 90% or more, or 95% or more of the total RNA present in the composition. In some aspects, the ssRNA vector comprises 97% or more, 98% or more, 99% or more, or 99.5% or more of the total RNA present in the composition. In some aspects, the ssRNA vector is the only RNA species present in the composition. In some aspects, the total RNA, dsRNA, and/or ssRNA content is assessed by dot blot or ELISA, optionally wherein the dot blot or ELISA assessment comprises detection with an antibody specific to RNA, dsRNA, and/or ssRNA. In some aspects, the total RNA, dsRNA, and/or ssRNA content is assessed by capillary electrophoresis. In some aspects, the total RNA, dsRNA, and/or ssRNA content is assessed liquid chromatography. In some aspects, the RNA, dsRNA, and/or ssRNA content is quantified as area under the curve (AUC).

[0013] In some aspects, the ssRNA vector is produced by *in vitro* transcription. In some aspects, the m⁷G cap comprises an m⁷G cap analog. In some aspects, the m⁷G cap analog comprises a trinucleotide m⁷G-ppp-A-U cap analog or dinucleotide m⁷G-ppp-A cap analog.

[0014] In some aspects, the ssRNA vector comprises a cassette comprising at least one nucleic acid sequence for delivery, optionally wherein the at least one nucleic acid sequence comprises a polypeptide-encoding nucleic acid sequence, optionally wherein the polypeptide-encoding nucleic acid sequence is an antigen-encoding nucleic acid sequence, and wherein the cassette is operably linked to or operably inserted into the self-amplifying backbone.

[0015] In some aspects, the self-amplifying backbone comprises a polynucleotide selected from a self-replicating RNA virus comprising at least one nucleic acid sequence, from 5' to 3', described by the formula:

$m^7G\text{-ppp-N}_1\text{-N}_2\text{-N}_v$, wherein

m^7G is a 7-methylguanylate (m^7G) cap,

ppp is a triphosphate bridge,

N_1 is a first nucleotide of the self-amplifying backbone corresponding to a first endogenous 5' nucleotide of the self-replicating RNA virus,

N_2 is a second nucleotide of the self-amplifying backbone corresponding to a second endogenous 5' nucleotide of the self-replicating RNA virus, and

N_v comprises (1) one or more additional nucleic acid sequences of the self-amplifying backbone, and (2) a cassette comprising at least one nucleic acid sequence for delivery, optionally wherein the at least one nucleic acid sequence comprises a polypeptide-encoding nucleic acid sequence, optionally wherein the polypeptide-encoding nucleic acid sequence is an antigen-encoding nucleic acid sequence, and wherein the cassette is operably linked to or operably inserted into the self-amplifying backbone.

[0016] In some aspects, the composition further comprising a nanoparticle delivery vehicle. In some aspects, the nanoparticle delivery vehicle is a lipid nanoparticle (LNP).

[0017] Also provided herein is a pharmaceutical composition comprising any one of the compositions described herein and a pharmaceutically acceptable carrier.

[0018] Also provided herein is a method for treating a subject with a disease, the method comprising administering to the subject an amount of any one of the compositions described herein or any one of the pharmaceutical compositions described herein. In some aspects, the disease is cancer or an infectious disease. In some aspects, the infectious disease is caused by a virus selected from the group comprising, HPV, influenza, TB, CMV, HMPV, PIV, chikungunya virus, Zika virus, SARS-CoV-2, and pan-coronavirus.

[0019] In some aspects, the amount of the ssRNA vector in the composition is 1000 μg or less, 100 μg or less, 50 μg or less, 30 μg or less, 10 μg or less, 5 μg or less, or 1 μg or less.

[0020] In some aspects, there is a reduction an innate immune response in the subject compared to a control subject receiving a composition comprising a nucleic acid sequence

comprising no modified nucleosides. In some aspects, the innate immune response is an IFN response. In some aspects, the IFN response is IRF-3 expression, IRF-7 expression, or a combination thereof.

[0021] In some aspects, there is an increase in replication of the ssRNA vector compared to a control composition comprising an otherwise identical ssRNA vector but comprising no modified nucleosides.

[0022] Also provided herein is a method of purifying a self-amplifying expression system from a nucleic acid mixture, the method comprising purifying the self-amplifying expression system through a cellulose chromatography system or an affinity-based separation system, wherein the self-amplifying expression system comprises a single-stranded RNA (ssRNA) vector, and wherein the ssRNA vector comprises an m⁷G cap, a polyA tail, and a self-amplifying backbone.

[0023] Also provided herein is a method of reducing double stranded RNA (dsRNA) in a nucleic acid mixture, the method comprising:

(A) purifying the self-amplifying expression system through a cellulose chromatography system or an affinity-based separation system; and/or

(B) producing the self-amplifying expression system such that the self-amplifying expression system comprises modified nucleosides; and

wherein the self-amplifying expression system comprises a single-stranded RNA (ssRNA) vector, wherein the ssRNA vector comprises an m⁷G cap, a polyA tail, and a self-amplifying backbone.

[0024] In some aspects, the affinity-based separation system comprises a deoxythymidine (dT) oligonucleotide (Oligo(dT)) system. In some aspects, the ssRNA vector comprises a modified nucleoside. In some aspects, the modified nucleoside comprises methyl-5-cytosine (m5C), methyl-6-adenosine (m6A), ribose-methylation (2'-O-Me), s-thiouridine, (s2U), 5-methyluridine (m5U), N1-methyl pseudouridine (m1Ψ), pseudouridine (Ψ), or any combination thereof. In some aspects, the modified nucleoside consists of methyl-5-cytosine (m5C).

[0025] Also provided herein is a composition mixture, wherein the composition comprises a self-amplifying expression system comprising:

(a) a single-stranded RNA (ssRNA) vector, wherein the ssRNA vector comprises an m⁷G cap, a polyadenylation (polyA) tail, and a self-amplifying backbone; and

(b) double stranded RNA (dsRNA), and

wherein the dsRNA comprises less than 20%, less than 10%, less than 5%, less than 2%, less than 1%, less than 0.5%, or less than 0.1% of the total RNA present in the mixture.

[0026] Also provided herein is a composition mixture, wherein the composition comprises a self-amplifying expression system comprising:

- (a) a single-stranded RNA (ssRNA) vector, wherein the ssRNA vector comprises an m⁷G cap, a polyadenylation (polyA) tail, and a self-amplifying backbone; and
- (b) double stranded RNA (dsRNA), and

wherein the ssRNA vector comprises 70% or more, 80% or more, 85% or more, 90% or more, 95% or more, 97% or more, 98% or more, 99% or more, or 99.5% or more of the total RNA present in the mixture.

[0027] In some aspects, the only detectable RNA comprises the ssRNA vector. In some aspects, the total RNA, dsRNA, and/or ssRNA content is assessed by dot blot or ELISA, optionally wherein the dot blot or ELISA assessment comprises detection with an antibody specific to RNA, dsRNA, and/or ssRNA. In some aspects, the total RNA, dsRNA, and/or ssRNA content is assessed by capillary electrophoresis. In some aspects, the total RNA, dsRNA, and/or ssRNA content is assessed liquid chromatography. In some aspects, the RNA, dsRNA, and/or ssRNA content is quantified as area under the curve (AUC).

[0028] In some aspects, the ssRNA is produced by *in vitro* transcription. In some aspects, the m⁷G cap comprises an m⁷G cap analog. In some aspects, the m⁷G cap analog comprises a trinucleotide m⁷G-ppp-A-U cap analog or dinucleotide m⁷G-ppp-A cap analog.

[0029] In some aspects, the ssRNA vector comprises a cassette comprising at least one nucleic acid sequence for delivery, optionally wherein the at least one nucleic acid sequence comprises a polypeptide-encoding nucleic acid sequence, optionally wherein the polypeptide-encoding nucleic acid sequence is an antigen-encoding nucleic acid sequence, and wherein the cassette is operably linked to or operably inserted into the self-amplifying backbone. In some aspects, the self-amplifying backbone comprises a polynucleotide selected from a self-replicating RNA virus comprising at least one nucleic acid sequence, from 5' to 3', described by the formula:

m⁷G-ppp-N₁-N₂-N_v, wherein

m⁷G is a 7-methylguanylate (m⁷G) cap,

ppp is a triphosphate bridge,

N₁ is a first nucleotide of the self-amplifying backbone corresponding to a first endogenous 5' nucleotide of the self-replicating RNA virus,

N₂ is a second nucleotide of the self-amplifying backbone corresponding to a second endogenous 5' nucleotide of the self-replicating RNA virus, and

N_v comprises (1) one or more additional nucleic acid sequences of the self-amplifying backbone, and (2) a cassette comprising at least one nucleic acid sequence for delivery,

optionally wherein the at least one nucleic acid sequence comprises a polypeptide-encoding nucleic acid sequence, optionally wherein the polypeptide-encoding nucleic acid sequence is an antigen-encoding nucleic acid sequence, and wherein the cassette is operably linked to or operably inserted into the self-amplifying backbone.

[0030] In some aspects, the ssRNA vector comprises one or more modified nucleosides. In some aspects, the one or more modified nucleosides comprise methyl-5-cytosine (m5C), methyl-6-adenosine (m6A), ribose-methylation (2'-O-Me), s-thiouridine, (s2U), 5-methyluridine (m5U), N1-methyl pseudouridine (m1Ψ), pseudouridine (Ψ), or any combination thereof. In some aspects, the one or more modified nucleosides comprise m5C. In some aspects, the one or more modified nucleosides comprise m5C and no other modified nucleotides, optionally other than a cap analog.

[0031] In some aspects, the ssRNA vector is purified. In some aspects, the ssRNA vector is purified by chromatography. In some aspects, the chromatography comprises a cellulose chromatography system or an affinity-based separation system. In some aspects, the affinity-based separation system is a deoxythymidine (dT) oligonucleotide (Oligo(dT)) system

[0032] In some aspects, the composition for delivery of the self-amplifying expression system comprises:

(A) the self-amplifying expression system, wherein the self-amplifying expression system comprises one or more self-amplifying mRNA (SAM) vectors, wherein the one or more SAM vectors comprise:

(a) the self-amplifying backbone, wherein the self-amplifying backbone comprises the nucleic acid sequence set forth in SEQ ID NO:6, wherein the self-amplifying backbone sequence comprises a subgenomic promoter nucleotide sequence and a poly(A) sequence, wherein the subgenomic promoter sequence is endogenous to the self-replicating RNA virus, wherein the poly(A) sequence is endogenous to the self-replicating RNA virus backbone; and

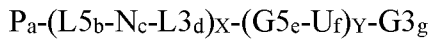
(b) the cassette integrated between the subgenomic promoter nucleotide sequence and the poly(A) sequence, wherein the cassette is operably linked to the subgenomic promoter nucleotide sequence, and optionally wherein the cassette comprises at least one antigen-encoding nucleic acid sequence comprising:

a. an epitope-encoding nucleic acid sequence, optionally comprising: (1) at least one alteration that makes the encoded epitope sequence distinct from the corresponding peptide sequence encoded by a wild-type nucleic acid sequence, or (2) 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) optionally, a lipid-nanoparticle (LNP), wherein the LNP encapsulates the self-amplifying expression system.

[0033] In some aspects, an ordered sequence of each element of the cassette in the composition for delivery of the self-amplifying expression system is described in the formula, from 5' to 3', comprising



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.

[0034] 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.

[0035] In some aspects, a = 0, b = 1, d = 1, e = 1, g = 1, h = 1, X = 10, Y = 2, the at least one promoter nucleotide sequence is a single subgenomic promoter nucleotide sequence provided by the self-amplifying backbone, the at least one polyadenylation poly(A) sequence is a poly(A) sequence of at least 80 consecutive A nucleotides provided by the self-amplifying backbone, the cassette is integrated between the subgenomic promoter nucleotide sequence and the poly(A) sequence, wherein the cassette is operably linked to the subgenomic 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 self-amplifying 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.

[0036] In some aspects, the at least one nucleic acid sequence for delivery comprises the polypeptide-encoding nucleic acid sequence. In some aspects, the polypeptide-encoding nucleic acid sequence encodes the antigen-encoding nucleic acid sequence. In some aspects, the antigen-encoding nucleic acid sequence comprises a MHC class I epitope, a MHC class II epitope, an epitope capable of stimulating a B cell response, or a combination thereof. In some aspects, the antigen-encoding nucleic acid sequence comprises sequence encoding a full-length protein, a protein subunit, a protein domain, or a combination thereof. In some aspects, the polypeptide-encoding nucleic acid sequence encodes a full-length protein or functional portion thereof. In some aspects, the full-length protein or functional portion thereof is selected from the group consisting of: an antibody, a cytokine, a chimeric antigen receptor (CAR), a T-cell receptor, and a genome-editing system nuclease.

[0037] In some aspects, the at least one nucleic acid sequence for delivery comprises at least one nucleic acid sequence comprising a non-coding nucleic acid sequence. In some aspects, the non-coding nucleic acid sequence is an RNA interference (RNAi) polynucleotide or genome-editing system polynucleotide.

[0038] 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 between 60-140nm.

[0039] In some aspects, the composition for delivery of the self-amplifying expression system is formulated for intramuscular (IM), intradermal (ID), subcutaneous (SC), intravitreal (IVT), intrathecal, or intravenous (IV) administration. In some aspects, the composition for delivery of the self-amplifying expression system is formulated for intramuscular (IM) administration.

[0040] 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.

[0041] In some aspects, the ssRNA vector comprises a positive-stranded RNA vector. In some aspects, the ssRNA vector comprises a negative-stranded RNA vector. In some aspects, the negative-stranded RNA vector comprises at least one polynucleotide sequence of a measles virus or a rhabdovirus. In some aspects, the ssRNA vector is self-amplifying within a mammalian cell. In some aspects, the self-replicating RNA virus is selected from the group consisting of: an alphavirus; a flavivirus, a measles, and a rhabdovirus. In some aspects, the self-amplifying backbone comprises at least one polynucleotide sequence of an alphavirus, optionally wherein the alphavirus is selected from the group consisting of: Aura virus, a Fort Morgan virus, a Venezuelan equine encephalitis virus, a Ross River virus, a Semliki Forest virus, a Sindbis virus, and a Mayaro virus. In some aspects, the self-amplifying backbone comprises at least one nucleotide sequence of a Venezuelan equine encephalitis virus. In some aspects, the self-amplifying backbone comprises at least sequences for nonstructural protein-mediated amplification, a subgenomic 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 self-amplifying backbone comprises at least sequences for nonstructural protein-mediated amplification, a subgenomic 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 self-amplifying backbone does not encode structural virion proteins capsid, E2 and E1, optionally wherein E1 is a full-length E1, or does not encode structural virion proteins Capsid, E3, E2, 6K. In some aspects, the cassette is inserted in place of structural virion proteins within the polynucleotide 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 self-amplifying 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 in SEQ ID NO:6 or SEQ ID NO:7. 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 for delivery are in separate open reading frames.

[0042] In some aspects, the ssRNA vector comprises at least one promoter nucleotide sequence. In some aspects, the at least one promoter nucleotide sequence comprises the native promoter nucleotide sequence encoded by the self-replicating RNA virus, optionally wherein the native promoter nucleotide sequence is a subgenomic promoter nucleotide sequence. In some aspects, the at least one promoter nucleotide sequence is an exogenous RNA promoter. In some aspects, the ssRNA vector comprises a second promoter sequence, optionally wherein the second promoter nucleotide sequence is a subgenomic promoter nucleotide sequence. In some aspects, the second promoter nucleotide sequence comprises multiple subgenomic promoter nucleotide sequences, wherein each subgenomic promoter nucleotide sequence provides for transcription of one or more of the separate open reading frames.

[0043] In some aspects, the ssRNA vector is at least 300nt in size. In some aspects, the ssRNA vector is at least 1kb in size. In some aspects, the ssRNA vector is at least 2kb in size. In some aspects, the ssRNA vector is less than 5kb in size.

[0044] 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 comprises 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.

[0045] 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 epitope-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.

[0046] 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, each antigen-encoding nucleic acid sequence independently comprises at least 2-10, 2, 3, 4, 5, 6, 7, 8, 9, or 10 epitope-encoding nucleic acid sequences, optionally wherein each epitope-encoding nucleic acid sequence encodes a distinct epitope-encoding nucleic acid sequence. In some aspects, each antigen-encoding nucleic acid sequence independently 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 epitope-encoding nucleic acid sequences, optionally wherein each epitope-encoding nucleic acid sequence encodes a distinct epitope-encoding nucleic acid sequence. In some aspects, each antigen-encoding nucleic acid sequence independently 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 epitope-encoding

nucleic acid sequences. In some aspects, each antigen-encoding nucleic acid sequence independently comprises at least 2-400 epitope-encoding nucleic acid sequences and wherein at least two of the epitope-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 a cell surface, optionally a tumor cell surface or an infected cell surface.

[0047] 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. 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.

[0048] 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.

[0049] 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.

[0050] 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 poly(A) sequence comprises a poly(A) sequence native to the self-replicating virus. In some aspects, the at least one poly(A) sequence comprises a poly(A) sequence exogenous to the self-replicating virus. 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, at least 90, at

least 100, at least 110, or at least 120 consecutive A nucleotides. In some aspects, the at least one poly(A) sequence is at least 80 consecutive A nucleotides.

[0051] In some aspects, the epitope-encoding nucleic acid sequence comprises an 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 a tumor, an infected cell, or an infectious disease organism, wherein the 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 a cell surface, optionally a tumor cell surface or an 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.

[0052] 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 nucleotide sequencing data from a tumor, an infected cell, or an infectious disease organism, wherein the 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 a cell surface, optionally a tumor cell surface or an 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.

[0053] 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 a cell surface, optionally a tumor cell surface or an 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.

[0054] In some aspects, selecting the set of selected epitopes comprises selecting epitopes that have an increased likelihood of being presented on a cell surface, optionally a tumor cell surface or an 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 stimulating a tumor-specific or 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 stimulating 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 a tumor cell or tissue, an infected cell, or an infectious disease organism. In some aspects, the sequencing is next generation sequencing (NGS) or any massively parallel sequencing approach.

[0055] In some aspects, the composition for delivery of the self-amplifying expression system is administered as a priming vaccine. In some aspects, the method further comprises administering a second composition, optionally wherein the second composition is a vaccine composition. In some aspects, the second composition is administered prior to the composition for delivery of the self-amplifying expression system. In some aspects, the second composition is administered subsequent to the administration of the composition for delivery of the self-amplifying expression system. In some aspects, the second composition is the same as the composition for delivery of the self-amplifying expression system. In some aspects, the second composition is different from the composition for delivery of the self-amplifying expression system. In some aspects, the second composition comprises the cassette of the self-amplifying expression system, optionally wherein the second composition comprises a chimpanzee adenovirus vector encoding the cassette of the self-amplifying expression system. In some

aspects, two or more second compositions are administered, optionally wherein the composition for delivery of the self-amplifying expression system is administered as a priming vaccine.

[0056] In some aspects, the composition for delivery of the self-amplifying expression system is administered intramuscularly (IM), intradermally (ID), subcutaneously (SC), intravitreal (IVT), intrathecal, or intravenously (IV). In some aspects, the method further comprises administering an immune modulator, optionally 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, an anti-OX-40 antibody or an antigen-binding fragment thereof, or a cytokine, optionally wherein the cytokine is at least one of IL-2, IL-7, IL-12, IL-15, or IL-21 or variants thereof. In some aspects, the method further comprises administering an adjuvant.

Brief Description of the Several Views of the Drawings

[0057] 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:

[0058] **Figure (FIG.) 1** illustrates transcription of SAM vectors using either a canonical T7 promoter or a modified (“minimal”) T7 promoter.

[0059] **FIG. 2** provides a schematic of a representative AU-SAM vector.

[0060] **FIG. 3** shows RNA profile of nucleoside-modified SAM encoding GFP or β -Spike by capillary electrophoresis. 3 ng of in vitro-transcribed SAM-GFP or β -Spike containing no or Ψ , m 1Ψ , and m5C-modified nucleosides were separated by capillary electrophoresis. U* SAM backbone contains catalytically inactive nsP4.

[0061] **FIG. 4** shows SAM activity measure by RT-qPCR following transfection. BHK-21 cells were transfected with 11 ng of indicated SAMs. RT-qPCR was performed with total RNA isolated 20 hours after transfection to measure the RNA copy number. Results are presented as mean +/- standard deviation for each group from 2 independent experiments.

[0062] **FIG. 5** shows FPLC-based Oligo(dT) affinity chromatography of in vitro transcribed SAM. A total of 5 mg or 11,760 nt long U-containing IVT SAM was purified by FPLC-based Oligo(dT) affinity Chromatography. **FIG. 5A**—A representative UV chromatogram of the affinity purification run with SAM encoding β -Spike is shown. The collected fractions containing IVT contaminants are marked as Flow through and Wash. The Oligo(dT)-bound full-length SAM eluted with water is marked as Elute. **FIG. 5B**—Analysis of the collected fractions by capillary electrophoresis.

[0063] FIG. 6 shows an evaluation of the purity of SAM by dot blot. dsRNA by-products in silica and Oligo(dT)-purified SAM-GFP or β -Spike containing no or m5C or m1 Ψ -modified nucleosides as analyzed by dot blotting with: (**FIG. 6A**) the J2 dsDNA-specific monoclonal antibody; and (**FIG. 6B**) relative quantitation of dsRNA content in the 1,000ng dot of each group.

[0064] FIG. 7A shows induction of innate immune response in SAM-transfected MoDCs. Human MoDCs were transfected with LNP-containing the indicated SAM (10 μ g) before Oligo(dT)-purification. Mean fluorescence intensity (MFI) of IRF7 was determined by flow-cytometry at 24 hours post-transfection. Results are presented as mean +/- standard deviation for each group from 4 independent experiments.

[0065] FIG. 7B shows induction of innate immune response in SAM-transfected MoDCs. Human MoDCs were transfected with LNP-containing the indicated SAM (10 μ g) after Oligo(dT)-purification. Mean fluorescence intensity (MFI) of IRF7 was determined by flow-cytometry at 24 hours post-transfection. Results are presented as mean +/- standard deviation for each group from 4 independent experiments.

[0066] FIG. 7C shows induction of innate immune response in SAM-transfected MoDCs. Human MoDCs were transfected with LNP-containing the indicated SAM (10 μ g) before (" β -Spike") and after purification (" β -Spike-ssRNA") using both oligo(dT)-mediated affinity purification (left panel) or cellulose-based purification (right panel). Shown is a representative flow-cytometry at 24 hours post-transfection.

[0067] FIG. 8 shows translation of SAM encoding GFP in MoDCs. Human MoDCs were transfected with LNP encapsulated SAM (10 μ g). GFP expression was determined by flow-cytometry at 48 hours post-transfection. Results are presented as mean +/- standard deviation for each group from 3-4 independent experiments.

[0068] FIG. 9 shows antigen-specific cellular immune response to SARS-CoV-2 β -Spike epitope in Balb/c mice at 6 or 14 days post-immunization with 1 or 10 μ g of SAM-LNP. Shown are IFN γ ELISpot responses. Cellulose-based purification is referred to as "Purified A" and oligo(dT)-based purification is referred to as "Purified B".

[0069] FIG. 10 shows intracellular cytokine staining in splenocytes of Balb/c mice following stimulation with overlapping peptide pool spanning SARS-CoV-2 β -Spike epitope in Balb/c mice at 5 or 12 days post-immunization with 1 mg of samRNA-LNP. Box represents interquartile range (IQR) (25 – 75%) with median line, whiskers represent range.

[0070] FIG. 11 shows serum pseudovirus neutralization titers (50% inhibition) in Balb/c mice 4 or 8 weeks post immunization with 1 mg of samRNA-LNP. Geomean and geometric SD of n = 6 independent animals per group.

[0071] FIG. 12 shows intracellular cytokine staining in splenocytes of Balb/c mice following stimulation with tetramer AH1 antigen in Balb/c mice at 3, 7, 10 & 14 days post-immunization with 1 mg of samRNA-LNP. Box represents interquartile range (IQR) (25 – 75%) with median line, whiskers represent range.

Detailed Description

Definitions

[0072] As used herein, the term "pharmaceutically acceptable salt" refers to those salts which are, within the scope of sound medical judgment, suitable for use in contact with the tissues of humans and lower animals without undue toxicity, irritation, allergic response and the like, and are commensurate with a reasonable benefit/risk ratio. Pharmaceutically acceptable salts are well known in the art. For example, S. M. Berge et al., describe pharmaceutically acceptable salts in detail in *J. Pharmaceutical Sciences*, 1977, 66, 1-19, incorporated herein by reference. Pharmaceutically acceptable salts of the compounds of this disclosure include those derived from suitable inorganic and organic acids and bases. Examples of pharmaceutically acceptable, nontoxic acid addition salts are salts of an amino group formed with inorganic acids such as hydrochloric acid, hydrobromic acid, phosphoric acid, sulfuric acid and perchloric acid or with organic acids such as acetic acid, oxalic acid, maleic acid, tartaric acid, citric acid, succinic acid or malonic acid or by using other methods used in the art such as ion exchange. Other pharmaceutically acceptable salts include adipate, alginate, ascorbate, aspartate, benzenesulfonate, benzoate, bisulfate, borate, butyrate, camphorate, camphorsulfonate, citrate, cyclopentanepropionate, digluconate, dodecylsulfate, ethanesulfonate, formate, fumarate, glucoheptonate, glycerophosphate, gluconate, hemisulfate, heptanoate, hexanoate, hydroiodide, 2-hydroxy-ethanesulfonate, lactobionate, lactate, laurate, lauryl sulfate, malate, maleate, malonate, methanesulfonate, 2-naphthalenesulfonate, nicotinate, nitrate, oleate, oxalate, palmitate, pamoate, pectinate, persulfate, 3-phenylpropionate, phosphate, pivalate, propionate, stearate, succinate, sulfate, tartrate, thiocyanate, p-toluenesulfonate, undecanoate, valerate salts, and the like.

[0073] Salts derived from appropriate bases include alkali metal, alkaline earth metal, ammonium and N(C1-4alkyl)₄ salts. Representative alkali or alkaline earth metal salts include sodium, lithium, potassium, calcium, magnesium, and the like. Further pharmaceutically acceptable salts include, when appropriate, nontoxic ammonium, quaternary ammonium, and amine cations formed using counterions such as halide, hydroxide, carboxylate, sulfate, phosphate, nitrate, loweralkyl sulfonate and aryl sulfonate.

[0074] The recitation of a listing of chemical groups in any definition of a variable herein includes definitions of that variable as any single group or combination of listed groups. The recitation of an embodiment for a variable herein includes that embodiment as any single embodiment or in combination with any other embodiments or portions thereof.

[0075] The term "biological sample", as used herein, includes, without limitation, cell cultures or extracts thereof; biopsied material obtained from a mammal or extracts thereof; and blood, saliva, urine, feces, semen, tears, or other body fluids or extracts thereof.

[0076] As used herein, a "therapeutically effective amount" means an amount of a substance (e.g., a therapeutic agent, composition, and/or formulation) that stimulates a desired biological response. In some embodiments, a therapeutically effective amount of a substance is an amount that is sufficient, when administered as part of a dosing regimen to a subject suffering from or susceptible to a disease, disorder, and/or condition, to treat, diagnose, prevent, and/or delay the onset of the disease, disorder, and/or condition. As will be appreciated by those of ordinary skill in this art, the effective amount of a substance may vary depending on such factors as the desired biological endpoint, the substance to be delivered, the target cell or tissue, etc. For example, the effective amount of a provided compound in a formulation to treat a disease, disorder, and/or condition is the amount that alleviates, ameliorates, relieves, inhibits, prevents, delays onset of, reduces severity of and/or reduces incidence of one or more symptoms or features of the disease, disorder, and/or condition. In some embodiments, a "therapeutically effective amount" is at least a minimal amount of a provided compound, or composition containing a provided compound, which is sufficient for treating one or more symptoms of a disease or disorder.

[0077] Disease, disorder, and condition are used interchangeably herein.

[0078] As used herein, the terms "treatment," "treat," and "treating" refer to partially or completely alleviating, inhibiting, delaying onset of, preventing, ameliorating and/or relieving a disorder or condition, or one or more symptoms of the disorder or condition, as described herein. In some embodiments, treatment may be administered after one or more symptoms have developed. In some embodiments, the term "treating" includes preventing or halting the progression of a disease or disorder. In other embodiments, treatment may be administered in the absence of symptoms. For example, treatment may be administered to a susceptible individual prior to the onset of symptoms (e.g., in light of a history of symptoms and/or in light of genetic or other susceptibility factors). Treatment may also be continued after symptoms have resolved, for example to prevent or delay their recurrence. Thus, in some embodiments, the term "treating" includes preventing relapse or recurrence of a disease or disorder.

[0079] A “subject” to which administration is contemplated includes, but is not limited to, humans (*i.e.*, a male or female of any age group, *e.g.*, a pediatric subject (*e.g.*, infant, child, adolescent) or adult subject (*e.g.*, young adult, middle-aged adult or senior adult)) and/or a non-human animal, *e.g.*, a mammal such as primates (*e.g.*, cynomolgus monkeys, rhesus monkeys), cattle, pigs, horses, sheep, goats, rodents, cats, and/or dogs. In certain embodiments, the subject is a human. In certain embodiments, the subject is a non-human animal. The terms “patient,” and “subject” are used interchangeably herein.

[0080] The term “pharmaceutically acceptable carrier, adjuvant, or vehicle” refers to a non-toxic carrier, adjuvant, or vehicle that does not destroy the pharmacological activity of the compound(s) with which it is formulated. Pharmaceutically acceptable carriers, adjuvants or vehicles that may be used in the compositions of the compounds disclosed herein include, but are not limited to, ion exchangers, alumina, aluminum stearate, lecithin, serum proteins, such as human serum albumin, buffer substances such as phosphates, glycine, sorbic acid, potassium sorbate, partial glyceride mixtures of saturated vegetable fatty acids, water, salts or electrolytes, such as protamine sulfate, disodium hydrogen phosphate, potassium hydrogen phosphate, sodium chloride, zinc salts, colloidal silica, magnesium trisilicate, polyvinyl pyrrolidone, cellulose-based substances, polyethylene glycol, sodium carboxymethylcellulose, polyacrylates, waxes, polyethylene-polyoxypropylene-block polymers, polyethylene glycol and wool fat.

Alternative Embodiments

[0081] In an alternative embodiment, compounds described herein may also comprise one or more isotopic substitutions. For example, hydrogen may be ^2H (D or deuterium) or ^3H (T or tritium); carbon may be, for example, ^{13}C or ^{14}C ; oxygen may be, for example, ^{18}O ; nitrogen may be, for example, ^{15}N , and the like. In other embodiments, a particular isotope (*e.g.*, ^3H , ^{13}C , ^{14}C , ^{18}O , or ^{15}N) can represent at least 1%, at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 99%, or at least 99.9% of the total isotopic abundance of an element that occupies a specific site of the compound.

Pharmaceutical Compositions

[0082] In some embodiments, the present disclosure provides a composition comprising any of the compositions described herein (*e.g.*, any of the ssRNA SAM vaccine compositions described herein) and a pharmaceutically acceptable carrier, adjuvant, or vehicle. Compounds of the present disclosure are preferably formulated in dosage unit form for ease of administration and uniformity of dosage.

Methods of Using Compounds of the Present Disclosure – Synthesis of RNA nucleotides

[0083] In general, the methods described herein may be useful in the preparation of *in vitro* transcribed RNA. In some embodiments, the methods described herein may be useful in the preparation of single-stranded RNA (ssRNA). In some embodiments, the methods described herein may be useful in the preparation of a self-amplifying expression system, such as self-amplifying expression systems that include ssRNA (*e.g.*, ssRNA vectors derived from self-replicating RNA viruses). In some embodiments, the methods described herein may be useful in the preparation of modified RNA, such as RNA (*e.g.*, ssRNA) incorporating modified nucleosides. In some embodiments, the methods described herein may be useful in the preparation of purified ssRNA, *e.g.*, ssRNA incorporating modified nucleosides.

[0084] In some embodiments, the methods described herein may be useful in the preparation of a 5'-capped RNA. Methods and compositions contemplated herein for preparation of 5'-capped RNA include, but are not limited to, mRNA, small nuclear RNA (snRNA), small nucleolar RNA (snoRNA), small cajal body-specific RNA (scaRNA). In some embodiments, a method involves the use of a cap containing oligonucleotide primers, nucleoside 5'-triphosphates (NTPs) and RNA polymerase for DNA-templated and promoter-controlled synthesis of RNA. In certain aspects, a method uses an initiating capped oligonucleotide primer that provides utility in RNA synthesis, in particular synthesis of capped mRNAs.

[0085] In some embodiments, the methods described herein may be useful in a method for preparation of RNA including, but not limited to, mRNA, snRNA, snoRNA, scaRNA, transfer RNA (tRNA), ribosomal RNA (rRNA), and transfer-messenger RNA (tmRNA) that carry modifications at or near 5'-end of the molecule. In some embodiments, a method involves the use of initiating oligonucleotide primers with or without Cap, nucleoside 5'-triphosphates (NTPs) and RNA polymerase for DNA-templated and promoter-controlled synthesis of RNA. In certain aspects, a method uses a modified initiating oligonucleotide primer carrying structural modifications that provide utility in RNA synthesis; in particular synthesis of 5'-modified RNAs.

[0086] The initiating capped oligonucleotide primer has an open 3'-OH group that allows for initiation of RNA polymerase mediated synthesis of RNA on a DNA template by adding nucleotide units to the 3'-end of the primer. The initiating capped oligonucleotide primer is substantially complementary to template DNA sequence at the transcription initiation site (*i.e.*, the initiation site is located closer to 3'-terminus of a promoter sequence and may overlap with promoter sequence), in certain embodiments, the initiating capped oligonucleotide primer directs synthesis of RNA predominantly in one direction ("forward") starting from the 3'-end of the primer. In certain aspects and embodiments, the initiating capped oligonucleotide primer outcompetes any nucleoside 5'-triphosphate for initiation of RNA synthesis, thereby maximizing

the production of the RNA that starts with initiating capped oligonucleotide primer and minimizing a production of RNA that starts with 5'- triphosphate-nucleoside (typically GTP).

[0087] An initiating capped oligonucleotide primers of the present disclosure have a hybridization sequence which may be complementary to a sequence on DNA template at initiation site. The presence of hybridization sequence forces an initiating capped oligonucleotide primer to predominantly align with complementary sequence of the DNA template at the initiation site in only the desired orientation (i.e., the "forward" orientation). In the forward orientation, the RNA transcript begins with the inverted guanosine residue (i.e., ^{7m}G(5')ppp(5') N...). The dominance of the forward orientation of the primer alignment on DNA template over incorrect "reverse" orientation is maintained by the thermodynamics of the hybridization complex. The latter is determined by the length of the hybridization sequence of initiating capped oligonucleotide primer and the identity of bases involved in hybridization with DNA template. Hybridization in the desired forward orientation may also depend on the temperature and reaction conditions at which DNA template and initiating capped oligonucleotide primer are hybridized or used during in vitro transcription.

[0088] An initiating capped oligonucleotide primer of the present disclosure enhances efficacy of initiation of transcription compared to efficacy of initiation with standard GTP, ATP, CTP or UTP. In some embodiments, initiation of transcription is considered enhanced when synthesis of RNA starts predominantly from initiating capped oligonucleotide primer and not from any NTP in transcription mixture. The enhanced efficiency of initiation of transcription results in a higher yield of RNA transcript. The enhanced efficiency of initiation of transcription may be increased to about 10%, about 20%, about 40%, about 60%, about 80%, about 90%, about 100%, about 150%, about 200% or about 500% over synthesis of RNA with conventional methods without initiating capped primer. In certain embodiments "initiating capped oligonucleotide primers" out-compete any NTP (including GTP) for initiation of transcription. One of ordinary skill in the art is able to readily determine the level of substrate activity and efficacy of initiating capped oligonucleotide primers. In certain embodiments, initiation takes place from the capped oligonucleotide primer rather than an NTP, which results in a higher level of capping of the transcribed mRNA.

[0089] In some aspects, methods are provided in which RNA is synthesized utilizing an initiating capped oligonucleotide primer that has substitutions or modifications. In some aspects, the substitutions and modifications of the initiating capped oligonucleotide primer do not substantially impair the synthesis of RNA. Routine test syntheses can be pre-formed to determine if desirable synthesis results can be obtained with the modified initiating capped oligonucleotide primers. Those skilled in the art can perform such routine experimentation to

determine if desirable results can be obtained. The substitution or modification of initiating capped oligonucleotide primer include for example, one or more modified nucleoside bases, one or more modified sugars, one or more modified inter-nucleotide linkage and/or one or more modified triphosphate bridges.

[0090] The modified initiating capped oligonucleotide primer, which may include one or more modification groups of the methods and compositions provided herein, can be elongated by RNA polymerase on DNA template by incorporation of NTP onto open 3'-OH group. The initiating capped oligonucleotide primer may include natural RNA and DNA nucleosides, modified nucleosides or nucleoside analogs. The initiating capped oligonucleotide primer may contain natural internucleotide phosphodiester linkages or modifications thereof, or combination thereof.

Methods of Using Compounds of the Present Disclosure – Methods of Treatment

[0091] In some embodiments, the present disclosure provides a method for treating or lessening the severity of a disease or condition (*e.g.*, cancer or an infection) in a patient comprising the step of administering to said patient an RNA polynucleotide. In some embodiments, the present disclosure provides a prophylactic method for treating a disease or condition, *e.g.*, prophylactic vaccination.

[0092] In general, the methods described herein may be useful in the preparation *in vitro* transcribed RNA for use in a pharmaceutical composition for treating or lessening the severity of a disease or condition. In some embodiments, the methods described herein may be useful in the preparation of single-stranded RNA (ssRNA) for use in a pharmaceutical composition. In some embodiments, the methods described herein may be useful in the preparation of a self-amplifying expression system for use in a pharmaceutical composition, such as self-amplifying expression systems that include ssRNA (*e.g.*, ssRNA vectors derived from self-replicating RNA viruses). In some embodiments, the methods described herein may be useful in the preparation of modified RNA for use in a pharmaceutical composition, such as RNA (*e.g.*, ssRNA) incorporating modified nucleosides. In some embodiments, the methods described herein may be useful in the preparation of purified ssRNA for use in a pharmaceutical composition, *e.g.*, ssRNA incorporating modified nucleosides.

[0093] In some embodiments, compounds and compositions, according to a method of the present disclosure, may be administered using any amount and any route of administration effective for treating or lessening the severity of cancer. In some embodiments, a cancer is selected from the group consisting of lung cancer, melanoma, breast cancer, ovarian cancer, prostate cancer, kidney cancer, gastric cancer, colon cancer, testicular cancer, head and neck cancer, pancreatic cancer, bladder cancer, brain cancer, B-cell lymphoma, acute myelogenous

leukemia, adult acute lymphoblastic leukemia, chronic myelogenous leukemia, chronic lymphocytic leukemia, T cell lymphocytic leukemia, non-small cell lung cancer, and small cell lung cancer.

[0094] In some embodiments, cancer is a solid tumor. In some embodiments, cancer is selected from the group consisting of: microsatellite stable-colorectal cancer (MSS-CRC), non-small cell lung cancer (NSCLC), pancreatic ductal adenocarcinoma (PDA), and gastroesophageal adenocarcinoma (GEA). In some embodiments, cancer is selected from the group consisting of: MSS-CRC, NSCLC, and PDA.

[0095] In some embodiments, an RNA polynucleotide composition described herein, such as any of the nucleoside-modified and/or purified RNA compositions described herein (*e.g.*, any of the ssRNA SAM vaccine compositions described herein), is administered to a patient with cancer selected from the group consisting of lung cancer, melanoma, breast cancer, ovarian cancer, prostate cancer, kidney cancer, gastric cancer, colon cancer, rectal cancer, testicular cancer, head and neck cancer, pancreatic cancer, bladder cancer, brain cancer, B-cell lymphoma, acute myelogenous leukemia, adult acute lymphoblastic leukemia, chronic myelogenous leukemia, chronic lymphocytic leukemia, myeloma, T cell lymphocytic leukemia, non-small cell lung cancer, and small cell lung cancer.

[0096] In some embodiments, any of the RNA polynucleotide compositions described herein, such as any of the nucleoside-modified and/or purified RNA compositions described herein (*e.g.*, any of the ssRNA SAM vaccine compositions described herein), is administered to a patient with an infection. In some embodiments, an infection is a viral infection, fungal, or a bacterial infection. In some embodiments, an infection is a viral infection. In some embodiments, a viral infection is an infection by a virus, wherein the virus is HPV. In some embodiments, a viral infection is an infection by a virus, wherein the virus is an influenza virus. In some embodiments, a viral infection is an infection by a virus, wherein the virus is a coronavirus (*e.g.*, Severe acute respiratory syndrome (SARS) 2002-associated species, SARS-CoV-2 isolates and sub-isolates, Middle East respiratory syndrome (MERS) 2012-associated species, and/or other coronavirus species). In some embodiments, any of the compositions described herein (*e.g.*, any of the ssRNA SAM vaccine compositions described herein) is administered to a subject with COVID-19. In some embodiments, any of the compositions described herein (*e.g.*, any of the ssRNA SAM vaccine compositions described herein) is administered to a subject at risk for infection with COVID-19. In some embodiments, any of the compositions described herein (*e.g.*, any of the ssRNA SAM vaccine compositions described herein) is administered to a subject as a prophylactic vaccination against COVID-19.

[0097] In some embodiments, the present disclosure relates to a method of contacting a biological sample with an RNA polynucleotide compositions described herein, such as any of the nucleoside-modified and/or purified RNA compositions described herein (*e.g.*, any of the ssRNA SAM vaccine compositions described herein).

[0098] In some embodiments, one or more additional therapeutic agents, may also be administered in combination with an RNA polynucleotide compositions described herein, such as any of the nucleoside-modified and/or purified RNA compositions described herein such as any of the nucleoside-modified and/or purified RNA compositions described herein (*e.g.*, any of the ssRNA SAM vaccine compositions described herein). In some embodiments, an RNA polynucleotide composition described herein and one or more additional therapeutic agents may be administered as part of a multiple dosage regime. In some embodiments an RNA polynucleotide composition described herein and one or more additional therapeutic agents may be administered may be administered simultaneously, sequentially or within a period of time. In some embodiments, an RNA polynucleotide composition described herein and one or more additional therapeutic agents may be administered within five hours of one another. In some embodiments, an RNA polynucleotide composition described herein and one or more additional therapeutic agents may be administered within 24 hours of one another. In some embodiments, an RNA polynucleotide composition described herein and one or more additional therapeutic agents may be administered within one week of one another.

Self-Amplifying mRNA Vectors

[0099] In general, all self-amplifying mRNA (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, 5' untranslated regions (UTRs), 3' UTRs, and/or a polyA sequence). A self-amplifying backbone can also include sequences for expression of subgenomic viral RNA (*e.g.*, a subgenomic promoter, such as 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 greater detail in Lundstrom

(Molecules. 2018 Dec 13;23(12). pii: E3310. doi: 10.3390/molecules23123310), herein incorporated by reference for all purposes.

Self-Amplifying Production *in vitro*

[00100] A 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 (*e.g.*, chemical and/or enzymatic synthesis), and polymerase chain reaction (PCR).

[00101] The DNA template contains an 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, SP6, or K11. Depending on the specific RNA polymerase promoter sequence chosen, additional 5' nucleotides can be 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_v for the production of desired sequence N will result in the mRNA sequence GG-N_v. In general, and without wishing to be bound by theory, T7 polymerase more efficiently transcribes RNA transcripts beginning with guanosine. However, additional 5' nucleotides may not be desired and/or may be detrimental. Accordingly, 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 endogenous (also referred to as "native" or "genomic") 5' sequence of the self-replicating virus from which the SAM vector is derived, referring to the native genomic sequence of the self-replicating virus (*e.g.*, having endogenous 5' VEEV nucleotides AU also referred to as "AU-SAM"). For example, a minimal T7 promoter can be referred to by the sequence TAATACGACTCACTATA (oriented 5'-3'; φ6.5 T7 promoter), in which an IVT reaction using the DNA template TAATACGACTCACTATAN₁N₂N_v for the production of desired sequence N will result in the mRNA sequence N₁N₂N_v. An alternative minimal T7 promoter can be referred to by the sequence TAATACGACTCACTATT (oriented 5'-3'; φ2.5 T7 promoter). Likewise, a minimal SP6 promoter referred to by the sequence ATTTAGGTGACACTATA can be used to generate transcripts without additional 5' nucleotides. Likewise, a minimal K11 promoter referred to by the sequence AATTAGGGCACACTATA 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).

[00102] 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 analogs during IVT. Cap analogs can include dinucleotide (m^7G -ppp-N) cap analogs or trinucleotide (m^7G -ppp- N_1 - N_2) cap analogs, where N represents a nucleotide or modified nucleotide (*e.g.*, ribonucleosides including, but not limited to, adenosine, guanosine, cytidine, and uridine). A modified nucleotide can include a modified adenosine, such as N6-methyladenosine 2'-OH-methylated. In an illustrative non-limiting example including a trinucleotide (m^7G -ppp- N_1 - N_2) cap analog, N_1 can be N6-methyladenosine 2'-OH-methylated. Cap analogs can include any of the structures or formulas described herein. Exemplary cap analogs and their use in IVT reactions are also described in greater detail 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 analog (m^7G -ppp-N-N) can be used. The trinucleotide cap analog 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 analog (m^7G -ppp-N).

[00103] 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.

Modified Nucleosides

[00104] *In vitro* transcription reactions for the production of SAM vectors can be supplemented with nucleoside triphosphates (ATP, CTP, UTP, GTP) having modified nucleosides. Without wishing to be bound by theory, RNA incorporating modified nucleosides can improve stability, reduce stimulating innate immune pathways, and/or improve viral protein-mediated RNA replication/amplification of the RNA vectors following transfection of target cells.

[00105] Modified nucleosides include, but are not limited to, hypoxanthine, inosine, 8-oxo-adenine, 7-substituted derivatives thereof, dihydrouracil, pseudouracil, 2-thiouracil, 4-thiouracil, 5-aminouracil, 5-(C1-C6)-alkyluracil, 5-methyluracil, 5-(C2-C6)-alkenyluracil, 5-(C2-C6)-alkynyluracil, 5-(hydroxymethyl)uracil, 5-chlorouracil, 5-fluorouracil, 5-bromouracil, 5-hydroxy cytosine, 5-(C1-C6)-alkylcytosine, 5-methylcytosine, 5-(C2-C6)-alkenylcytosine, 5-(C2-C6)alkynylcytosine, 5-chlorocytosine, 5-fluorocytosine, 5-bromocytosine, N2-

dimethylguanine, 7-deazaguanine, 8-azaguanine, 7-deaza-7-substituted guanine, 7-deaza-7-(C2-C6)alkynylguanine, 7-deaza-8-substituted guanine, 8-hydroxyguanine, 6-thioguanine, 8-oxoguanine, 2-aminopurine, 2-amino-6-chloropurine, 2,4-diaminopurine, 2,6-diaminopurine, 8-azapurine, substituted 7-deazapurine, 7-deaza-7-substituted purine, 7-deaza-8-substituted purine, hydrogen (abasic residue). Modified nucleotides can include methyl-5-cytosine (m5C), methyl-6-adenosine (m6A), ribose-methylation (2'-O-Me), s-thiouridine, (s2U), 5-methyluridine (m5U), or N1-methyl pseudouridine (m1Ψ), pseudouridine (Ψ). Modified nucleotides can include m5C. Modified nucleotides can include m6A. Modified nucleotides can include 2'-O-Me. Modified nucleotides can include s-thiouridine. Modified nucleotides can include m5U. Modified nucleotides can include m1Ψ. Modified nucleotides can include Ψ.

[00106] *In vitro* transcription reactions can be supplemented with a single distinct modified nucleoside together with unmodified nucleosides. For example, as an illustrative non-limiting example, an *in vitro* transcription reactions can be supplemented only with m5C modified nucleosides and no other modified nucleosides (*e.g.*, together with unmodified ATP, UTP, and GTP). *In vitro* transcription reactions can be supplemented with multiple modified nucleosides. *In vitro* transcription reactions can be supplemented with multiple modified nucleosides together with unmodified nucleosides. Unmodified nucleosides can be inclusive or exclusive of m⁷G cap analogs that include chemical modifications.

[00107] An ssRNA incorporating modified nucleosides can have greater than 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, or 100% of adenosine, guanosine, cytidine, and/or uridine nucleosides of the ssRNA be modified nucleosides. An ssRNA incorporating modified nucleosides can have greater than 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, or 100% of adenosine nucleosides of the ssRNA be modified nucleosides. An ssRNA incorporating modified nucleosides can have greater than 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, or 100% of guanosine nucleosides of the ssRNA be modified nucleosides. An ssRNA incorporating modified nucleosides can have greater than 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, or 100% of cytidine nucleosides of the ssRNA be modified nucleosides. An ssRNA incorporating modified nucleosides can have greater than 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, or 100% of uridine nucleosides of the ssRNA be modified nucleosides.

[00108] An ssRNA incorporating modified nucleosides can have greater than 25% of adenosine nucleosides of the ssRNA be modified nucleosides. An ssRNA incorporating modified nucleosides can have greater than 50% of adenosine nucleosides of the ssRNA be modified nucleosides. An ssRNA incorporating modified nucleosides can have greater than 75% of adenosine nucleosides of the ssRNA be modified nucleosides. An ssRNA incorporating

modified nucleosides can have greater than 95% of adenosine nucleosides of the ssRNA be modified nucleosides. An ssRNA incorporating modified nucleosides can have greater than 99% of adenosine nucleosides of the ssRNA be modified nucleosides. An ssRNA incorporating modified nucleosides can have 100% of adenosine nucleosides of the ssRNA be modified nucleosides.

[00109] An ssRNA incorporating modified nucleosides can have greater than 25% of guanosine nucleosides of the ssRNA be modified nucleosides. An ssRNA incorporating modified nucleosides can have greater than 50% of guanosine nucleosides of the ssRNA be modified nucleosides. An ssRNA incorporating modified nucleosides can have greater than 75% of guanosine nucleosides of the ssRNA be modified nucleosides. An ssRNA incorporating modified nucleosides can have greater than 95% of guanosine nucleosides of the ssRNA be modified nucleosides. An ssRNA incorporating modified nucleosides can have greater than 99% of guanosine nucleosides of the ssRNA be modified nucleosides. An ssRNA incorporating modified nucleosides can have 100% of guanosine nucleosides of the ssRNA be modified nucleosides.

[00110] An ssRNA incorporating modified nucleosides can have greater than 25% of cytidine nucleosides of the ssRNA be modified nucleosides. An ssRNA incorporating modified nucleosides can have greater than 50% of cytidine nucleosides of the ssRNA be modified nucleosides. An ssRNA incorporating modified nucleosides can have greater than 75% of cytidine nucleosides of the ssRNA be modified nucleosides. An ssRNA incorporating modified nucleosides can have greater than 95% of cytidine nucleosides of the ssRNA be modified nucleosides. An ssRNA incorporating modified nucleosides can have greater than 99% of cytidine nucleosides of the ssRNA be modified nucleosides. An ssRNA incorporating modified nucleosides can have 100% of cytidine nucleosides of the ssRNA be modified nucleosides.

[00111] An ssRNA incorporating modified nucleosides can have greater than 25% of uridine nucleosides of the ssRNA be modified nucleosides. An ssRNA incorporating modified nucleosides can have greater than 50% of uridine nucleosides of the ssRNA be modified nucleosides. An ssRNA incorporating modified nucleosides can have greater than 75% of uridine nucleosides of the ssRNA be modified nucleosides. An ssRNA incorporating modified nucleosides can have greater than 95% of uridine nucleosides of the ssRNA be modified nucleosides. An ssRNA incorporating modified nucleosides can have greater than 99% of uridine nucleosides of the ssRNA be modified nucleosides. An ssRNA incorporating modified nucleosides can have 100% of uridine nucleosides of the ssRNA be modified nucleosides.

[00112] An ssRNA incorporating modified nucleosides can have greater than 25% of cytidine nucleosides of the ssRNA be m5c. An ssRNA incorporating modified nucleosides can have

greater than 50% of cytidine nucleosides of the ssRNA be m5c. An ssRNA incorporating modified nucleosides can have greater than 75% of cytidine nucleosides of the ssRNA be m5c. An ssRNA incorporating modified nucleosides can have greater than 95% of cytidine nucleosides of the ssRNA be m5c. An ssRNA incorporating modified nucleosides can have greater than 99% of cytidine nucleosides of the ssRNA be m5c. An ssRNA incorporating modified nucleosides can have 100% of cytidine nucleosides of the ssRNA be m5c.

ssRNA Purification

[00113] Following *in vitro* transcription of SAM vectors, with or without modified nucleosides, RNA can be purified. Without wishing to be bound by theory, the desired SAM vector is a single-stranded RNA (ssRNA) inclusive of an m⁷G cap and a polyadenylation (polyA) tail corresponding the full-length construct encoded by a production vector. Further, without wishing to be bound by theory, purification removes all undesired double-stranded RNA byproducts of *in vitro* transcription, including those that may result in undesired immune stimulation and/or reduced viral protein-mediated RNA replication/amplification of the RNA vectors following transfection of target cells. For example, stimulation of the innate immune system by RNA IVT products can be assessed, such as by assessing an IFN response. IFN responses can be assessed through analysis of Interferon expression, IRF-3 expression, IRF-7 expression, or a combination thereof (*e.g.*, by RT-qPCR, flow-cytometry, Western blot analysis, or a combination thereof). Improvements in replication of the SAM ssRNA vector can be assessed, such as through comparing SAM ssRNA that (A) include modified nucleosides; and/or (B) has been purified using a chromatography system and/or an affinity-based separation system relative to a control composition comprising an otherwise identical ssRNA vector but comprising no modified nucleosides and/or is unpurified.

[00114] Purification can use chromatography-based systems, such as FPLC or HPLC systems. In general, FPLC will be advantageous for the typical SAM ssRNA vector sizes described herein. One skilled in the art will recognize applicable FPLC systems and HPLC systems that can be employed, such as the Äkta avant 25 FPLC system (Cytiva), and how to use the same (*e.g.*, washes, equilibrations, elutions, etc.).

[00115] Chromatography-based systems include cellulose-based chromatography system. Chromatography-based systems include affinity-based separation systems. An illustrative non-limiting example of an affinity-based separation system is an deoxythymidine (dT) oligonucleotide (Oligo(dT))-based system that binds a polyA tail of RNA, such as a POROS OligodT(25) GoPure Column (Thermo Fisher Scientific).

[00116] Purification strategies, such as the chromatography-based strategies, can be used to produce pure, or essential pure, ssRNA free of contaminants, such as IVT reaction byproducts (*e.g.*, dsRNA, salt, detergent, etc.). Purification strategies, such as the chromatography-based strategies, can be used to produce a SAM ssRNA-containing composition or mixture, where a contaminant is less than 20%, less than 10%, less than 5%, less than 2%, less than 1%, less than 0.5%, or less than 0.1% of the composition or mixture. Purification strategies can be used to produce a mixture including both SAM ssRNA as well as dsRNA, but where the dsRNA is less than 20%, less than 10%, less than 5%, less than 2%, less than 1%, less than 0.5%, or less than 0.1% of the composition or mixture.

[00117] Purification strategies can be used to produce a SAM ssRNA-containing composition or mixture, where the desired ssRNA vector (*e.g.*, a full-length ssRNA vector including both an m⁷G cap and polyA tail) comprises 70% or more, 80% or more, 85% or more, 90% or more, or 95% or more of the total RNA present in the composition or mixture. Purification strategies can be used to produce a SAM ssRNA-containing composition or mixture, where the desired ssRNA vector comprises 97% or more, 98% or more, 99% or more, or 99.5% or more of the total RNA present in the composition of the composition or mixture.

[00118] Purification strategies can be used to produce a mixture including both SAM ssRNA as well as dsRNA, but where the ssRNA is 70% or more, 80% or more, 85% or more, 90% or more, or 95% or more of the RNA present in the mixture. Purification strategies can be used to produce a mixture including both SAM ssRNA as well as dsRNA, but where the ssRNA is 97% or more, 98% or more, 99% or more, or 99.5% or more of the RNA present in the mixture.

[00119] Purification strategies can be used to produce a SAM ssRNA-containing composition or mixture, where the desired ssRNA vector is the only detectable RNA species present in the composition of the composition or mixture. Purification strategies can be used to produce a SAM ssRNA-containing composition or mixture, where the desired ssRNA vector is the only RNA species present in the composition of the composition or mixture.

[00120] Purity of ssRNA following purification, such as following a chromatography-based strategy, can be assessed. RNA (*e.g.*, total RNA, dsRNA, and/or ssRNA) quality and concentration can be assessed using methods including, but not limited to, capillary electrophoresis, liquid chromatography (*e.g.*, FPLC or HPLC), dot blot, and/or ELISA. Dot blots and ELISAs can use detection methods that include an antibody specific to RNA, dsRNA, and/or ssRNA. Quantification can include determining an area under the curve (AUC), such as FPLC spectral peaks, band intensity, and/or dot intensity.

[00121] RNA purification processes can also include other purification techniques well-known in the field, such as phenol-chloroform extraction.

Alphavirus Biology

[00122] 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).

[00123] 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.

[00124] 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.

Alphavirus Delivery Vector

[00125] 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 mRNA (SAM) 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 stimulate 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 VEEV. Thus, an example of an antigen expression vector described herein can utilize an alphavirus backbone that allows for a high level of antigen expression, stimulates a robust immune response to antigen, does not stimulate an immune response to the vector itself, and can be used in a safe manner. Furthermore, the antigen expression cassette can be designed to stimulate different levels of an immune response through optimization of which alphavirus sequences the vector uses, including, but not limited to, sequences derived from VEEV or its attenuated derivative TC-83.

[00126] Several expression vector design strategies have been engineered using alphavirus sequences (Pushko 1997). In one strategy, an 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.

[00127] 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 express 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.

Delivery Via Lipid Nanoparticles (LNP)

[00128] 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.

[00129] 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.

[00130] 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.

[00131] 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. As illustrative examples, selective and targeted delivery of LNP can be achieved by 1) incorporating lipid conjugated ligands (e.g., mannose) to cell-type specific receptors into LNP and/or 2) incorporating into LNP a membrane-tethering lipoprotein (Anchor) that interacts with the targeting antibodies. The anchor can be protein A/G and any structural form of antibodies including scFv, Fab, and VHH single domain antibody or nanobodies with extrinsic lipidation signal (e.g., palmitoylation, prenylation, and miristoylation) encoded either at its N-terminus or at its C-terminus. 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, phosphocholine, phosphoethanolamine, a sterol, or neutral lipids.

[00132] 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.

Other Vectors

[00133] Self-amplifying mRNA (SAM) based compositions described herein can be used together with other compositions featuring distinct (*e.g.*, non-SAM) vector backbones. For example SAM compositions can be used as part of a vaccine strategy that also uses vector backbones of chimpanzee origin to encode an antigen cassette. 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 are described further in USPN 6,083,716, US Application Pub. No. US20200197500A1, and international patent application publication WO2020/243719, each of which is herein incorporated by reference in its entirety, for all purposes.

Antigens

[00134] 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.

[00135] Disclosed herein are isolated peptides that comprise tumor specific mutations identified by the methods disclosed herein, peptides that comprise known tumor specific mutations, and mutant polypeptides or fragments thereof identified by methods disclosed herein. Neoantigen peptides can be described in the context of their coding sequence where a neoantigen includes the nucleotide sequence (*e.g.*, DNA or RNA) that codes for the related polypeptide sequence.

[00136] Also disclosed herein are peptides derived from any polypeptide known to or have been found to have altered expression in a tumor cell or cancerous tissue in comparison to a normal cell or tissue, for example any polypeptide known to or have been found to be aberrantly expressed in a tumor cell or cancerous tissue in comparison to a normal cell or tissue. Suitable polypeptides from which the antigenic peptides can be derived can be found for example in the COSMIC database. COSMIC curates comprehensive information on somatic mutations in human cancer. A peptide can contain a tumor specific mutation. Tumor antigens (*e.g.*, shared tumor antigens and tumor neoantigens) can include, but are not limited to, those described in US App. No. 17/058,128, herein incorporated by reference for all purposes.

[00137] 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 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 includes, 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.

[00138] 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.

[00139] Vectors and associated compositions described herein can be used to deliver antigens from any organism, including their toxins or other by-products, to prevent and/or treat infection or other adverse reactions associated with the organism or its by-product.

[00140] 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 stimulating 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 hemorrhagic 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 cultivivirus (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. 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 alpha herpesvirinae, which encompasses the genera simplexvirus (HSV I, HSV II), varicellovirus (pseudorabies, varicella zoster) and the sub-family beta herpesvirinae, which includes the genera cytomegalovirus (Human CMV), muromegalovirus) and the sub-family gamma herpesvirinae, which includes the genera lymphocryptovirus, EBV (Burkitt's 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.

[00141] 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 sommus*); *moraxella*; *H. ducreyi* (which causes chancroid); *brucella*; *Francisella tularensis* (which causes tularemia); *yersinia (pasteurella)*; *streptobacillus moniliformis* and *spirillum*. Gram-positive bacilli include *listeria monocytogenes*; *erysipelothrix 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.

[00142] 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.

[00143] Antigens can be selected that are predicted to be presented on the cell surface of a cell, such as a tumor cell, 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.

[00144] 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.

[00145] One or more antigens can be presented on the surface of a tumor. One or more antigens can be presented on the surface of an infected cell.

[00146] One or more antigens can be immunogenic in a subject having a tumor, e.g., capable of stimulating a T cell response and/or a B cell response in the subject. One or more antigens can be immunogenic in a subject having or suspected to have an infection, e.g., capable of stimulating 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, e.g., capable of stimulating 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.

[00147] One or more antigens can be capable of stimulating 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 antigen). 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. Antigens capable of stimulating a B cell response to an infection can be an antigen found on the surface of an infectious disease organism. Antigens capable of eliciting a B cell response to an infection can be an intracellular antigen expressed in an infectious disease organism.

[00148] One or more antigens can include a combination of antigens capable of stimulating a T cell response (*e.g.*, peptides including predicted T cell epitope sequences) and distinct antigens capable of stimulating a B cell response (*e.g.*, full-length proteins, protein subunits, protein domains).

[00149] One or more antigens that stimulate an autoimmune response in a subject can be excluded from consideration in the context of vaccine generation for a subject.

[00150] 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.

[00151] 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.

[00152] 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) neoepitope sequence present in the tumor (*e.g.* due to a frameshift, read-through or intron inclusion that leads to a novel peptide sequence), a longer peptide would consist of: (3) the entire stretch of novel tumor-specific or 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 stimulation of T cell responses. Longer peptides can also include a full-length protein, a protein subunit, a protein domain, and combinations thereof of a peptide, such as those expressed in an infectious disease organism. Longer peptides (*e.g.*, full-length protein, protein subunit, or protein domain) and combinations thereof can be included to stimulate a B cell response.

[00153] 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.

[00154] In some aspects, antigenic peptides and polypeptides do not stimulate an autoimmune response and/or invoke immunological tolerance when administered to a subject.

[00155] 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 varies by length, amino acid sequence, or both. Tumor-specific peptides can be derived from any polypeptide known to or have been found to contain a tumor specific mutation or peptides derived from any polypeptide known to or have been found to have altered expression in a tumor cell or cancerous tissue in comparison to a normal cell or tissue, for example any polypeptide known to or have been found to be aberrantly expressed in a tumor cell or cancerous tissue in comparison to a normal cell or tissue. 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). Suitable polypeptides from which the antigenic peptides can be derived can be found for example in the COSMIC database or the AACR Genomics Evidence Neoplasia Information Exchange (GENIE) database. COSMIC curates comprehensive information on somatic mutations in human cancer. AACR GENIE aggregates and links clinical-grade cancer genomic data with clinical outcomes from tens of thousands of cancer patients. A peptide can include a tumor-specific mutation. In some aspects the tumor specific mutation is a driver mutation for a particular cancer type.

[00156] 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).

[00157] 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.

[00158] 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 stimulate CTL activity can be enhanced by linkage to a sequence which contains at least one epitope that is capable of stimulating 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.

[00159] An 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.

[00160] 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.

[00161] 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. A 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.

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.

Vaccine Compositions

[00162] Also disclosed herein is an immunogenic composition, e.g., a vaccine composition, capable of raising a specific immune response, e.g., a tumor-specific immune response or an 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.

[00163] 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.

[00164] 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 in greater detail 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).

[00165] 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.

[00166] A vaccine can contain at least two repeats of an epitope-encoding nucleic acid sequence. As 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

[00167] 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.

[00168] 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:

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

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.

[00169] 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 and/or a neoantigen 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.

[00170] 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.*, $E_A-E_A\dots$ 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.*, $E_A-E_B-E_C-E_A\dots$, 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 $E_A-E_B-E_A\dots$, E_A is separated by 75 nucleotides. In an illustrative example, an antigen-encoding nucleic acid having the sequence

VTNTEMFVTAPDNLGYMYEVQWPGQTQPQIANCSVYDFFVWLHYYSVRDTVTNTEMFVTAPDNLGYMYEVQWPGQTQPQIANCSVYDFFVWLHYYSVRDT 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.

[00171] 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.

[00172] The vaccine composition can be capable of stimulating a specific cytotoxic T-cell response and/or a specific helper T-cell response. The vaccine composition can be capable of stimulating a specific cytotoxic T-cell response and a specific helper T-cell response.

[00173] The vaccine composition can be capable of stimulating a specific B-cell response (*e.g.*, an antibody response).

[00174] The vaccine composition can be capable of stimulating a specific cytotoxic T-cell response, a specific helper T-cell response, and/or a specific B-cell response. The vaccine composition can be capable of stimulating a specific cytotoxic T-cell response and a specific B-cell response. The vaccine composition can be capable of stimulating a specific helper T-cell response and a specific B-cell response. The vaccine composition can be capable of stimulating a specific cytotoxic T-cell response, a specific helper T-cell response, and a specific B-cell response.

[00175] 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, *e.g.*, a dendritic cell (DC) capable of presenting the peptide to a T-cell.

[00176] 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.

[00177] 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.

[00178] 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).

[00179] 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.

[00180] 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).

[00181] 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.

[00182] 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.

[00183] 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.

[00184] 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 stimulate 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.

Antigen Cassette

[00185] The methods employed for the selection of one or more antigens, the cloning and construction of an “antigen cassette” and its insertion into a viral vector are within the skill in the art given the teachings provided herein. By "antigen 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. A cassette can include one or more antigens, such as one or more pathogen-derived peptides, virus-derived peptides, bacteria-derived peptides, fungus-derived peptides, parasite-derived peptides, and/or tumor-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).

[00186] 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.

[00187] 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.

[00188] An 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.

[00189] As described elsewhere herein, the antigen cassette can be located in the site of any selected deletion in a viral vector, such as the deleted structural proteins of a VEEV backbone or

the site of the E1 gene region deletion or E3 gene region deletion of a ChAd-based vector, among others which may be selected.

[00190] The antigen cassette can be described using the following formula to describe the ordered sequence of each element, from 5' to 3':

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

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 N_c is a epitope encoding nucleic acid sequence, where for each Y the corresponding U_f is a universal MHC class II epitope-encoding nucleic acid sequence. A universal sequence can comprise at least one of Tetanus toxoid and PADRE. A universal sequence can comprise a Tetanus toxoid peptide. A universal sequence can comprise a PADRE peptide. A universal sequence can comprise a Tetanus toxoid and PADRE peptides. 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.

[00191] 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 a vector backbone, such as an RNA alphavirus backbone, is present), 10 MHC class I epitopes 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 a vector backbone (*e.g.*, an RNA alphavirus backbone). Examples of linking the 3' end of the antigen cassette to a vector backbone (*e.g.*, an RNA alphavirus backbone) include linking directly to the 3' UTR elements provided by the vector backbone, such as a 3' 19-nt CSE. Examples of linking the 5' end of the antigen cassette to a vector backbone (*e.g.*, an RNA alphavirus backbone) include linking directly to a promoter or 5' UTR element of the vector backbone, such as subgenomic promoter sequence (*e.g.*, a 26S subgenomic promoter sequence), an alphavirus 5' UTR, a 51-nt CSE, or a 24-nt CSE.

[00192] Other examples include: where $a = 1$ describing where a promoter other than the promoter nucleotide sequence provided by a vector backbone (*e.g.*, an RNA alphavirus 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 a vector backbone (*e.g.*, an RNA alphavirus backbone).

[00193] 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.

[00194] 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.

[00195] 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.

[00196] The promoter nucleotide sequences P and/or P2 can be the same as a promoter nucleotide sequence provided by a vector backbone, such as an RNA alphavirus backbone. For example, the promoter sequence provided by the vector backbone, P_n and P2, can each comprise a subgenomic promoter sequence (*e.g.*, 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 a vector backbone (*e.g.*, an RNA alphavirus backbone), as well as can be different from each other.

[00197] The 5' linker L5 can be a native sequence or a non-natural sequence. Non-natural sequence includes, 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.

[00198] 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 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.

[00199] 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 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.

[00200] For each X, each N can encode a MHC class I epitope, a MHC class II epitope, an epitope/antigen capable of stimulating a B cell response, or a combination thereof. For each X, each N can encode a combination of a MHC class I epitope, a MHC class II epitope, and an epitope/antigen capable of stimulating a B cell response. For each X, each N can encode a

combination of a MHC class I epitope and a MHC class II epitope. For each X, each N can encode a combination of a MHC class I epitope and an epitope/antigen capable of stimulating a B cell response. For each X, each N can encode a combination of a MHC class II epitope and an epitope/antigen capable of stimulating a B cell response. For each X, each N can encode a MHC class II epitope. For each X, each N can encode an epitope/antigen capable of stimulating a B cell response. For each X, each N can encode an MHC class I epitope 7-15 amino acids in length. For each X, each N can also encode an 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 encode an 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.

[00201] 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 infectious disease or tumor derived nucleic acid sequences 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.

[00202] 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.

[00203] 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.

[00204] 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.

Immune Modulators

[00205] 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. An immune modulator can include a binding molecule (e.g., an antibody such as an scFv) which binds to and blocks the activity of an immune checkpoint molecule. An immune modulator can include a cytokine, such as IL-2, IL-7, IL-12 (including IL-12 p35, p40, p70, and/or p70-fusion constructs), IL-15, or IL-21. An immune modulator can include a modified cytokine (e.g., pegIL-2). Vectors can comprise an antigen cassette and one or more nucleic acid molecules encoding an immune modulator.

[00206] 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.

Payload-Encoding SAM Compositions

[00207] Also disclosed herein is a SAM vector having the endogenous 5' sequence of the self-replicating virus from which the SAM vector is derived (*e.g.*, having endogenous 5' VEEV nucleotides AU also referred to as "AU-SAM") encoding one or more payload nucleic acid sequences, such as in a cassette. By "cassette" is meant the combination of a selected polynucleotide(s) (*e.g.*, antigen-encoding nucleic acid sequences) and the other regulatory elements necessary to transcribe the polynucleotide (s) and, generally in instances of coding sequences, express the transcribed product. Also disclosed herein is a SAM vector delivery composition capable of delivering one or more payload nucleic acid sequences. A payload nucleic acid sequence can be any nucleic acid sequence desired to be delivered to a cell of interest. In general, the payload is a nucleic acid sequence linked to a promoter or any translational tools (*e.g.*, IRES, any 2A self-cleaving peptide sequences such as P2A, E2A, F2A, and T2A) to drive expression of the nucleic acid sequence. The payload nucleic acid sequence can encode a polypeptide (*i.e.*, a nucleic acid sequence capable of being transcribed and translated into a protein). In general, a payload nucleic acid sequence encoding a peptide can encode any protein desired to be expressed in a cell. Examples of proteins include, but are not

limited to, an antigen (*e.g.*, an MHC class I epitope, an MHC class II epitope, or an epitope capable of stimulating a B cell response), an antibody, a cytokine, a chimeric antigen receptor (CAR), a T-cell receptor, or a genome-editing system component (*e.g.*, a nuclease used in a genome-editing system). Genome-editing systems include, but are not limited to, a CRISPR system, a zinc-finger system, a meganuclease system, or a TALEN system. The payload nucleic acid sequence can be non-coding (*i.e.*, a nucleic acid sequence capable of being transcribed but is not translated into a protein). In general, a non-coding payload nucleic acid sequence can be any non-coding polynucleotide desired to be expressed in a cell. Examples of non-coding polynucleotides include, but are not limited to, RNA interference (RNAi) polynucleotides (*e.g.*, antisense oligonucleotides, shRNAs, siRNAs, miRNAs etc.) or genome-editing system polynucleotide (*e.g.*, a guide RNA [gRNA] with various/different lengths, a single-guide RNA [sgRNA], a trans-activating CRISPR [tracrRNA], and/or a CRISPR RNA [crRNA]). A payload nucleic acid sequence can encode two or more (*e.g.*, 2, 3, 4, 5 or more) distinct polypeptides (*e.g.*, two or more distinct epitope sequences linked together) or contain two or more distinct non-coding nucleic acid sequences (*e.g.*, two or more distinct RNAi polynucleotides). A payload nucleic acid sequence can have a combination of polypeptide-encoding nucleic acid sequences and non-coding nucleic acid sequences.

Antigen Identification

[00208] 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, and international patent application publications WO/2018/195357 and WO/2018/208856, each herein incorporated by reference, in their entirety, for all purposes.

[00209] Methods for identifying antigens (*e.g.*, antigens derived from a tumor or an infectious disease organism) include identifying antigens that are likely to be presented on a cell surface (*e.g.*, presented by MHC on a tumor cell, 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 a tumor, an infected cell, or an infectious disease organism, wherein the nucleotide sequencing data and/or expression data is used to obtain data representing peptide sequences of each of a set of antigens (*e.g.*, antigens derived from the tumor or 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 a cell surface, such as a tumor cell or an 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.

[00210] Truncal peptides, meaning those presented by all or most subclones, can be prioritized for inclusion into a vaccine. 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 a vaccine.

[00211] 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 a tumor, an infectious disease, and/or 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 tumor or infectious disease escape)

[00212] 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 tumor or 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.

Therapeutic and Manufacturing Methods

[00213] Also provided is a method of stimulating a tumor specific immune response in a subject, vaccinating against a tumor, treating and/or alleviating a symptom of cancer in a subject by administering to the subject one or more antigens such as a plurality of antigens identified using methods disclosed herein.

[00214] Also provided is a method of stimulating an infectious disease organism-specific immune response in a subject, vaccinating against an 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.

[00215] In some aspects, a subject has been diagnosed with cancer or is at risk of developing cancer. A subject can be a human, dog, cat, horse or any animal in which a tumor specific immune response is desired. A tumor can be any solid tumor such as breast, ovarian, prostate, lung, kidney, gastric, colon, testicular, head and neck, pancreas, brain, melanoma, and other tumors of tissue organs and hematological tumors, such as lymphomas and leukemias, including acute myelogenous leukemia, chronic myelogenous leukemia, chronic lymphocytic leukemia, T cell lymphocytic leukemia, and B cell lymphomas.

[00216] 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.

[00217] An antigen can be administered in an amount sufficient to stimulate a CTL response. An antigen can be administered in an amount sufficient to stimulate a T cell response. An antigen can be administered in an amount sufficient to stimulate a B cell response.

[00218] An 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.

[00219] 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. The optimum amount of a checkpoint inhibitor dosing regimen can be determined. For example, a checkpoint inhibitor composition 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.

[00220] 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.

[00221] A vaccine can be compiled so that the selection, number and/or amount of antigens present in the composition is/are tissue, cancer, 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 specific type of cancer, the specific infectious disease (*e.g.* a specific infectious disease isolate/strain 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.

[00222] 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.

[00223] For a composition to be used as a vaccine for cancer or 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 tumor or infected cell of a patient expresses high amounts of a certain antigen, the respective pharmaceutical composition for treatment of this cancer or 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.

[00224] Compositions comprising an antigen can be administered to an individual already suffering from cancer or an infection. In therapeutic applications, compositions are administered to a patient in an amount sufficient to stimulate an effective CTL response to the tumor antigen or 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 a cancer has metastasized or an 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.

[00225] For therapeutic use, administration can begin at the detection or surgical removal of tumors, or 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).

[00226] Pharmaceutical compositions (*e.g.*, vaccine compositions) for therapeutic treatment are intended for parenteral, topical, nasal, oral or local administration. Pharmaceutical compositions can be administered parenterally, *e.g.*, intravenously, subcutaneously, intradermally, or intramuscularly. The compositions can be administered at a site of surgical

excision to stimulate a local immune response to a tumor. 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.

[00227] 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.

[00228] 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.

[00229] 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.

[00230] 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).

[00231] 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 stimulate 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.

[00232] 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.

[00233] 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.

[00234] Also disclosed is a method of manufacturing a vaccine, comprising performing the steps of a method disclosed herein; and producing a vaccine comprising a plurality of antigens or a subset of the plurality of antigens.

[00235] 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.

[00236] 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.

Antigen Use and Administration

[00237] 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.

[00238] A priming vaccine can be based on SAM vaccine compositions described herein with a SAM having an endogenous 5' sequence of the self-replicating virus from which the SAM vector is derived (*e.g.*, endogenous 5' VEEV nucleotides AU also referred to as "AU-SAM").

[00239] A boosting vaccine (including two or more boosting administrations) can be based on SAM vaccine compositions described herein with a SAM having an endogenous 5' sequence of the self-replicating virus from which the SAM vector is derived (*e.g.*, endogenous 5' VEEV nucleotides AU also referred to as "AU-SAM").

[00240] A vaccination protocol can include both a priming vaccine and a boosting vaccine each based on SAM vaccine compositions described herein with a SAM having an endogenous 5' sequence of the self-replicating virus from which the SAM vector is derived (*e.g.*, endogenous 5' VEEV nucleotides AU also referred to as "AU-SAM").

[00241] A priming vaccine, including for use in combination with a SAM having an endogenous 5' sequence, can also be based on C68 (*e.g.*, the sequences shown in SEQ ID NO:1 or 2) or SAM (*e.g.*, the sequences shown in SEQ ID NO:3 or 4). A boosting vaccine, including for use in combination with a SAM having an endogenous 5' sequence, can also be based on C68 (*e.g.*, the sequences shown in SEQ ID NO:1 or 2) or SAM (*e.g.*, the sequences shown in SEQ ID NO:3 or 4).

[00242] Each vector in a prime/boost strategy typically includes a cassette that includes antigens. Cassettes can include about 1-50 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 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) an immune modulator. 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. CPI's can be administered via intravenous (i.v.) injection, sub-cutaneous (s.c.) injection, intradermal (i.d.) injection, intraperitoneal (i.p.) injection, intramuscular (i.m.) injection. Each vaccine dose can be administered to the subject in conjunction with (e.g., concurrently, before, or after) a cytokine, such as IL-2, IL-7, IL-12 (including IL-12 p35, p40, p70, and/or p70-fusion constructs), IL-15, or IL-21. Each vaccine dose can be administered to the subject in conjunction with (e.g., concurrently, before, or after) a modified cytokine (e.g., pegIL-2).

[00243] A priming vaccine can be injected (e.g., intramuscularly) in a subject. Uni- or 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 SAM vectors at low vaccine dose selected from the range 0.001 to 1 ug RNA.

[00244] 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. Uni- or 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 SAM vectors at low vaccine dose selected from the range 0.001 to 1 ug RNA.

[00245] A SAM dose can be administered as either as priming dose or as one or more boosting doses. Priming doses and boosting doses can be the same amount of SAM or different amounts of SAM. Each boosting dose can be the same amount of SAM or different amounts of SAM. A SAM dose between 10-30 μ g, 10-100 μ g, 10-300 μ g, 30-100 μ g, 30-300 μ g, or 100-300 μ g RNA can be administered. A SAM dose between 10-500 μ g, 10-1000 μ g, 30-500 μ g, 30-1000 μ g, or 500-1000 μ g RNA can be administered. A SAM dose between 1-30 μ g, 1-100 μ g, 1-300 μ g RNA can be administered. A SAM dose between 1-500 μ g or 1-1000 μ g RNA can be administered. A SAM dose of at least 400 μ g, at least 500 μ g, at least 600 μ g, at least 700 μ g, at least 800 μ g, at least 900 μ g, at least 1000 μ g RNA can be administered. A SAM dose of at least 1 μ g, 3 μ g, 10 μ g, or 30 μ g RNA can be administered. A SAM dose of 1 μ g, 3 μ g, 10 μ g, or 30 μ g RNA can be administered. A SAM dose between 1-3 μ g, 1-10 μ g, 1-30 μ g, 3-10 μ g, 3-30 μ g, or 10-

300µg RNA can be administered. A SAM dose of 10µg, 30µg, 100µg, or 300µg RNA can be administered. A SAM dose of 300µg RNA can be administered. A SAM dose of 100µg RNA can be administered. A SAM dose of 30µg RNA can be administered. A SAM dose of 10µg RNA can be administered. A SAM dose of 3µg RNA can be administered. A SAM dose of 1µg RNA can be administered. A SAM dose at least 300µg RNA can be administered. A SAM dose at least 100µg RNA can be administered. A SAM dose at least 30µg RNA can be administered. A SAM dose at least 10µg RNA can be administered. A SAM dose of at least 3µg RNA can be administered. A SAM dose of at least 1µg RNA can be administered. A SAM dose less than or equal to 300µg RNA can be administered.

[00246] 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 SAM 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.

[00247] 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.

[00248] Immune monitoring can be performed before, during, and/or after vaccine administration. Such monitoring can inform safety and efficacy, among other parameters.

[00249] 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).

[00250] Immune responses, 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). 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.

[00251] 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). Antibodies can also be assessed for function, such as assessed for neutralizing ability.

Exemplification

[00252] In order that the present disclosure described herein may be more fully understood, the following examples are set forth. The synthetic and biological examples described in this application are offered to illustrate the compounds, pharmaceutical compositions, and methods provided herein and are not to be construed in any way as limiting their scope.

Materials and Methods

[00253] The compounds provided herein can be prepared from readily available starting materials using the following general methods and procedures. It will be appreciated that where typical or preferred process conditions (*i.e.*, reaction temperatures, times, mole ratios of reactants, solvents, pressures, *etc.*) are given, other process conditions can also be used unless otherwise stated. Optimum reaction conditions may vary with the particular reactants or solvent used, but such conditions can be determined by one skilled in the art by routine optimization.

[00254] Additionally, as will be apparent to those skilled in the art, conventional protecting groups may be necessary to prevent certain functional groups from undergoing undesired reactions. The choice of a suitable protecting group for a particular functional group as well as suitable conditions for protection and deprotection are well known in the art. For example, numerous protecting groups, and their introduction and removal, are described in T. W. Greene and P. G. M. Wuts, *Protecting Groups in Organic Synthesis*, Second Edition, Wiley, New York, 1991, and references cited therein.

[00255] The compounds provided herein may be isolated and purified by known standard procedures. Such procedures include (but are not limited to) trituration, column chromatography, HPLC, or supercritical fluid chromatography (SFC). The compounds provided herein may be prepared from known or commercially available starting materials and reagents by one skilled in the art of organic synthesis.

Example 1. Self-amplifying Expression Systems

A. Self-replicating RNA Virus Backbone and SAM Generation

[00256] In one implementation of the present invention, an RNA alphavirus backbone for the expression system was generated from a self-replicating Venezuelan Equine Encephalitis virus (“VEEV”; Kinney, 1986, *Virology* 152: 400-413) by deleting the structural proteins of VEEV located 3' of the 26S subgenomic promoter, except the last 50 amino acids of E1 (VEEV sequences 7544 to 11,176 deleted; numbering based on Kinney *et al* 1986; SEQ ID NO:6). To generate the self-amplifying mRNA (“SAM”) ssRNA vector, the deleted sequences were replaced by payload sequences. A representative SAM vector containing 20 model antigens is “VEE-MAG25mer” (SEQ ID NO:4). A modified T7 RNA polymerase promoter

(TAATACGACTCACTATA), which lacks the canonical 3' dinucleotide GG, was added to the 5' end of the SAM vector to generate the *in vitro* transcription template DNA (SEQ ID NO:57; 7544 to 11,176 deleted without an inserted cassette). An additional template production vector was produced adding a PCR forward primer sequence and 3' restriction sites (SEQ ID NO:58; 7544 to 11,176 deleted without an inserted cassette).

[00257] RNA produced using the template above contains an m⁷G cap directly linked to the endogenous 5' VEEV nucleotide sequence, *i.e.*, no additional intervening nucleotides are present between the m⁷G cap and the endogenous 5' VEEV nucleotide sequence, such as the dinucleotide GG typically present when a canonical T7 RNA polymerase is used. ssRNA production of SAM vectors with backbones beginning with endogenous nucleotides AUG and using a canonical or modified (“minimal”) T7 promoter is illustrated in **FIG. 1**. SAM vectors without additional intervening nucleotides located between the m⁷G cap and the endogenous 5' AU nucleotides are referred to herein as “AU-SAM” vectors. A schematic of a representative AU-SAM vector is shown in **FIG. 2**.

[00258] Capped AU-SAM ssRNA vectors, containing cassettes encoding representative payload sequences (*e.g.*, GFP or β-Spike), were produced co-transcriptionally using the following steps:

- A DNA template was produced cloning an antigen cassette of interest into the *in vitro* transcription template DNA (SEQ ID NO:57)
- Capped RNA was produced by *in vitro* transcription (IVT), as outline below:
 - o Reaction contained: 1x transcription buffer (40 mM Tris (pH 7.9), 10 mM dithiothreitol, 2 mM spermidine, 0.002% Triton X-100, and 27 mM magnesium chloride) using final concentrations of 1x T7 RNA polymerase mix (Cat. No. E2040S; New England Biolabs); 0.025 mg/mL DNA transcription template (linearized by restriction digest or PCR amplified); 8 mM trinucleotide m⁷G-ppp-A-U cap analog (CleanCap Reagent AU; Cat. No. N-7114; TriLink) and 10 mmol/l final concentration each of adenine triphosphate (ATP), cytidine triphosphate (CTP), guanine triphosphate (GTP), and uridine triphosphate (UTP) (HiScribe T7 Quick High Yield RNA Synthesis Kit; New England Biolabs)
 - o SAMs with modified nucleosides, the transcription reaction was assembled with the replacement of one or two nucleotide triphosphate with the corresponding triphosphate derivative of the following modified nucleosides: 5-methylcytidine (m5C), 5-methoxycytidine (mo5C), pseudouridine (Ψ), N1-methylpseudouridine (m1Ψ), 5-methyluridine (m5U), 5-methoxyuridine (mo5U), 2-thiouridine (s2U), or 6-methyladenine (m6A) (TriLink).

- IVT Reaction conditions: Transcription reactions were incubated at 37°C for 2 hours and treated with final 2 U DNase I (Cat. No. AM2239; Thermo Fisher Scientific) /0.001 mg DNA transcription template in DNase I buffer for 1 hr at 37°C.
- Capped/modified AU-SAM was purified using Turbo DNase (Thermo Fisher Scientific) digestion followed by RNeasy Maxi Kit (Cat. No. 75162; Qiagen) or liquid chromatography.

[00259] Model payloads encoding green fluorescent protein (GFP) and β -Spike from SARS-CoV-2 variant were inserted into the deleted region of the VEEV backbone. Capped AU-SAM RNA was produced using a trinucleotide m⁷G-ppp-A-U cap analog, as described above.

[00260] The quantity and quality of the SAMs were determined with a Nanodrop spectrophotometer (Thermo Fisher Scientific). All RNA samples were analyzed by capillary electrophoresis for RNA structural integrity assurance and a representation of all SAMs.

[00261] All RNA samples were analyzed by capillary electrophoresis for RNA structural integrity assurance and a representation of all SAMs used in the study (**FIG. 3**). In a series of transcription reactions using protein-encoding plasmids and an RNA polymerase, full-length transcripts containing Ψ , m1 Ψ , and m5C were obtained. Modified SAMs analyzed by capillary electrophoresis were indistinguishable from their non-modified counterparts in such that all were intact and migrated as expected based on their sizes.

B. Replication activity of SAM by RT-qPCR

[00262] The replication activity of nucleoside-modified SAMs encoding green fluorescent protein (GFP) was measured by RT-qPCR following transfection into BHK-21 cells. BHK-21 cells were transfected with 11 ng of SAM complexed with Lipofectamine MessengerMAX transfection reagent (Cat. No. LMRNA008; Thermo Fisher Scientific) and total RNA was then isolated using RNeasy Micro Kit (Cat. No. 74004; Qiagen) according to the manufacturer's protocol. To measure the replication activity of SAM (*i.e.*, self-amplification of RNA), the isolated RNAs were reverse-transcribed and analyzed by qPCR to measure the RNA copy number of GFP transcripts at 20 hours post transfection. SAM copy number was relative to a SAM control containing a catalytically inactive polymerase subunit, nsP4, that was incapable of self-amplification. As shown in **FIG. 4**, while all SAMs with modified nucleosides can replicate, m5C alone-containing SAM ("GFP-m5C") replicated more efficiently than its unmodified counterpart ("GFP-U"), resulting in 24% more transcripts. The GFP-m5C also replicated more efficiently than Ψ , m1 Ψ , or m5C+ m1 Ψ -containing SAM.

C. Affinity purification of SAM

[00263] The applicability of affinity chromatographic approaches to synthetic SAM ssRNA vectors, which are 3 to 4 times longer than typical non-amplifying mRNAs, was unknown. Therefore, the capacity of affinity chromatographic purification methods to remove byproducts generated during in vitro transcription (IVT) (contaminants), decrease innate immunity, and/or improve the translation and vaccination efficacy of synthetic SAMs was investigated. Removal of IVT contaminants (*e.g.*, salt, detergents, and/or dsRNA) from SAM was performed using POROS OligodT(25) GoPure Column (Thermo Fisher Scientific) and a chromatography buffer containing 10 mM HEPES (pH 7.3), 1 mM EDTA, and 100 mM NaCl (buffer B) or 500 mM NaCl (buffer A). The column was connected to an Äkta avant 25 FPLC system (Cytiva) and equilibrated with 5-column volumes of buffer A at a flow rate of 2.5 mL/min. SAM sample in buffer A was then loaded to the column at a flow rate of 1.5 mL/min and washed with 3-column volumes of buffer A followed by 5-column volumes of buffer B at a flow rate of 2.5 mL/min. The chromatography buffer was changed to nuclease-free water to release the Oligo(dT)-bound SAM at a flow rate of 1.5 mL/min. The UV absorbance of the flowthrough and wash fraction was monitored at 260 nm, and the peak fraction corresponding to the eluted SAM was collected. For further analysis, the nucleic acid from the collected chromatographic peaks was recovered by adding 0.1 volume of 3 M NaOAc (pH 5.5) and 1 volume of isopropanol, followed by precipitation.

[00264] The FPLC chromatogram of SAM encoding β -Spike demonstrated a major peak eluted with water (**FIG. 5A**), which was collected and identified as the expected SAM product using capillary electrophoresis (**FIG. 5B**, “Elute”). SAMs with or without nucleoside modification encoding different sequences yielded similar patterns with varying relative heights for the preceding and succeeding peaks. Additional UV-absorbing products with shorter and longer retention times relative to the main SAM product were also observed.

[00265] RNA quality and purification efficiency was assessed by capillary electrophoresis (3ng total RNA loaded per lane and analyzed on an Agilent Fragment Analyzer). As shown in **FIG. 5B**, bands or smears below the expected ssRNA vector product were observed. However, as quantitated below in **Table A**, when the SAM purity was compared to the input SAM (“ β -Spike Input”), the oligo(dT)-purified ssRNA vector (“Elute”) demonstrated a 20% increase in the purity of the expected product as measured by the area under the curve (AUC) between nucleotide length 8500 and 13000nt compared to the total RNA signal (100nt-13000nt).

Table A - Capillary electrophoresis RNA quantification

Sample ID	Range	% Total
β -Spike Input	100 nt to 8499 nt	35.1 %
β -Spike Input	8500 nt to 13000 nt	63.7 %
β -Spike Elute	100 nt to 8499 nt	22.7 %
β -Spike Elute	8500 nt to 13000 nt	76.8 %

D. The effect of nucleoside modification and affinity purification on innate immunogenicity of SAM

[00266] A dot blot assay with J2 monoclonal antibody (mAb) that recognize dsRNA was used to determine whether in vitro-transcribed SAM contains dsRNA. Testing β -Spike and GFP-encoding in vitro transcripts containing either no nucleoside modifications, m1 Ψ , or m5C nucleoside modifications, all samples contained sequence-dependent dsRNA contamination (**FIG. 6A**). Unpurified IVT m1 Ψ , or m5C SAM contained notably less dsRNA contaminant than the corresponding U-containing SAM (**FIG. 6B**), suggesting that m1 Ψ or m5C suppressed dsRNA byproduct formation. Affinity purification of both unmodified and nucleoside modified SAM further reduced staining by dsRNA-specific mAb (**FIG. 6B**).

[00267] Human monocyte-derived dendritic cells (MoDCs) were used to measure innate immune responses, particularly RNA-sensing pathways. IRF7 (an activation marker regulating Interferon transcription) expression level was assessed by flow cytometry at 24 hours post transfection of SAM ssRNA vectors encapsulated in LNP. Consistent with the results of dsRNA quantitation by dot blot (**FIG. 6B**), FACS analysis demonstrated SAM containing m1 Ψ or m5C alone (**FIG. 7A**), or each in combination with affinity purification (**FIG. 7B**), induced 54-77% less IRF7 expression compared to unmodified SAM. The m1 Ψ modification reduced RNA-mediated MoDC activation more potently than m5C modifications, likely due to the observed reduced SAM replication activity (**FIG. 4**) generating less dsRNA intermediates. Oligo(dT)-mediated affinity purification or cellulose-based purification alone of SAM ssRNA vectors encoding β -Spike also induced ~70% less IRF7 expression relative to LNP-mediated transfection with unpurified SAM (**FIG. 7C**, left and right panels, respectively).

E. Translational efficacy of nucleoside modified SAM purified by Oligo(dT) affinity purification

[00268] FACS analyses of MoDCs at 48 hours following LNP-mediated SAM delivery demonstrated that m5C-modified SAM combined with Oligo(dT) purification noticeably improved translational capacity of SAM-GFP as compared to the corresponding control, or

m1Ψ-containing transcripts (**FIG. 8**). Not only did more cells express GFP when m5C-modified SAM was used, but also the cells exhibited more intense staining, suggesting higher cellular levels of GFP. The level of GFP staining in both number of positive cell and intensity was proportional to the amounts of delivered SAM-GFP added to MoDCs (data not shown).

F. Self-amplifying mRNA Virus Vector Evaluation in Mice

[00269] Efficacy of self-amplifying expression systems that include ssRNA vectors (SAM) encoding β-Spike having (A) modified nucleosides; and/or (B) having been purified a chromatography system and/or an affinity-based separation system, as described herein, was assessed in mice. Efficacy was assessed through monitoring T cell responses.

Immunizations

[00270] Balb/c mice (n = 12 per group) were immunized with SAM-LNP. The SAM included modified nucleosides and/or chromatography purified SAM ssRNA. SAM-LNP complexes (1 μg and 10 μg) were administered in 100 μL volume, bilateral intramuscular injection (50 μL per leg). The immunization arms are shown below in **Table B** (Cellulose-based purification is referred to as “Purified A” and oligo(dT)-based purification is referred to as “Purified B”).

Table B – *In vivo* assessment arms of SAM with modified nucleosides and/or chromatography purified SAM ssRNA (Balb/c mice)

Group	Prime @ Day 0	Dose	Readout
1	SAM-β Spike	10 μg	T cells @ Days 6 & 14: ELISPOT and ICS with S1 and S2 antigens
2	SAM-β Spike	10 μg	
3	Cellulose-purified SAM-β Spike	10 μg	
4	Cellulose-purified SAM-β Spike	10 μg	
5	Cellulose-purified SAM-β Spike-m1Ψ	10 μg	
6	Cellulose-purified SAM-β Spike-m1Ψ	10 μg	
7	Cellulose-purified SAM-β Spike-m5C	10 μg	
8	Cellulose-purified SAM-β Spike-m1Ψ	10 μg	
9	Oligo(dT)-purified SAM-β Spike	10 μg	
10	Oligo(dT)-purified SAM-β Spike	10 μg	

Splenocyte dissociation

[00271] Splenocytes were isolated 6 days (n=6) and 12 days (n=6) post-immunization. Spleens for each mouse were pooled in 3 mL of complete RPMI (RPMI, 10% FBS, penicillin/streptomycin). Mechanical dissociation was performed using the gentleMACS

Dissociator (Miltenyi Biotec), following manufacturer's protocol. Dissociated cells were filtered through a 40 micron filter and red blood cells were lysed with ACK lysis buffer (150mM NH₄Cl, 10mM KHCO₃, 0.1mM Na₂EDTA). Cells were filtered again through a 30 micron filter and then resuspended in complete RPMI. Cells were counted on the Cytoflex LX (Beckman Coulter) using propidium iodide staining to exclude dead and apoptotic cells. Cell were then adjusted to the appropriate concentration of live cells for subsequent analysis.

Ex vivo enzyme-linked immunospot (ELISpot) analysis

[00272] ELISPOT analysis was performed according to ELISPOT harmonization guidelines {DOI: 10.1038/nprot.2015.068} with the mouse IFN γ ELISpotPLUS kit (MABTECH). 5x10⁴ splenocytes were stimulated *ex vivo* with 10uM of β -Spike peptides for 16 hours in 96-well IFN γ 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: spot count + 2 x (spot count x %confluence / [100% - %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.

G. Immunogenicity Results in Mice

[00273] Antigen-specific cellular immune responses to SARS-CoV-2 β -Spike epitope was measured in Balb/c mice at 6 or 14 days post-immunization with 1 or 10 μ g of SAM-LNP. As shown in **FIG. 9**, immune responses measured by IFN γ ELISpot displayed a trend in the epitope-specific T cell response suggesting that immunization with oligo(dT)-purified SAM improved T cell responses relative to the other SAM compositions assessed.

[00274] Antigen-specific cellular immune responses to SARS-CoV-2 β -Spike epitope was further measured in Balb/c mice post-immunization with 1mg of SAM-LNP produced and purified either with a standard "traditional" protocol or produced with m5C and purified with oligo(dT) ("Oligo(dT)-purified m5C- β -Spike samRNA"). **FIG. 10** shows immune responses measured by intracellular cytokine staining against overlapping peptide pools spanning SARS-CoV-2 β -Spike epitope in Balb/c mice at 5 or 12 days post-immunization. **FIG. 11** shows

pseudovirus neutralization titers (50% inhibition) in Balb/c mice 4 or 8 weeks post immunization.

[00275] Antigen-specific cellular immune responses to a model antigen cassette was further measured in Balb/c mice post-immunization with 1mg of SAM-LNP produced and purified either with a standard “traditional” protocol; produced with m5C and purified with oligo(dT) (“Oligo(dT)-purified m5C-MAG samRNA”); or produced with m1Ψ and purified with oligo(dT) (“Oligo(dT)-purified m1Ψ-MAG samRNA”). **FIG. 12** shows immune responses measured by intracellular cytokine staining in splenocytes of Balb/c mice following stimulation with tetramer AH1 antigen.

[00276] The results demonstrate samRNA produced with m5C and purified with oligo(dT) or produced with m1Ψ and purified with oligo(dT) stimulated immune responses.

H. Vaccine Efficacy Evaluation in Non-Human Primates

[00277] Efficacy and safety of vaccines containing cassettes encoding immunogens are evaluated in non-human primates. Specifically, self-amplifying expression systems that include ssRNA vectors (SAM) that (A) includes modified nucleosides; and/or (B) has been purified using a chromatography system and/or an affinity-based separation system, as described herein, are assessed. Efficacy is assessed through monitoring T cell and/or B cell responses.

Immunizations

[00278] For SAM vaccines in Indian rhesus macaques (*e.g.*, Mamu-A*01), SAM is administered as bilateral intramuscular injections into the quadriceps muscle, including but not limited to administering at a dose of 1 mg total per animal in 1 mL per leg.

Immune Monitoring in Rhesus

[00279] For immune monitoring, 10-20 mL of blood is collected into vacutainer tubes containing heparin and maintained at room temperature until isolation. PBMCs are isolated by density gradient centrifugation using lymphocyte separation medium (LSM) and Leucosep separator tubes. PBMCs are stained with propidium iodide and viable cells counted using the Cytoflex LX (Beckman Coulter). Samples are then resuspended at 4×10^6 cells/mL in RPMI complete (10% FBS).

[00280] IFN γ ELISPOT assays are performed using pre-coated 96-well plates (MABtech, Monkey IFN γ ELISPOT PLUS, ALP (Kit Lot #36, Plate Lot #19)) following manufacturer’s protocol. For each sample and stimuli, 1×10^5 PBMCs per well are plated in triplicate with 10 ug/mL peptide stimuli (GenScript) and incubated overnight in complete RPMI. Samples are incubated overnight with 10uM of respective peptides encoded by the SAM ssRNA vector.

DMSO only is used as a negative control for each sample. Plates are washed with PBS and incubated with anti-monkey IFN γ MAb biotin (MABtech) for two hours, followed by an additional wash and incubation with Streptavidin-ALP (MABtech) for one hour. After final wash, plates are incubated for ten minutes with BCIP/NBT (MABtech) to develop the immunospots and dried overnight at 37°C. Spots are imaged and enumerated using AID reader (Autoimmun Diagnostika).

[00281] Samples with replicate well variability (Variability = Variance/[median + 1]) greater than 10 and median greater than 10 are excluded. Spot values are adjusted based on the well saturation according to the formula: AdjustedSpots = RawSpots + 2*(RawSpots*Saturation/[100-Saturation]). Wells with well saturation greater than 33% are considered “too numerous to count” (TNTC) and excluded. Background correction for each sample is performed by subtracting the average value of the negative control peptide wells. Data is normalized to spot forming colonies (SFC) per 1×10^6 PBMCs by multiplying the corrected spot number by 1×10^6 /Cell number plated. For overall summary analysis calculated values generated by plating cells at 1×10^5 cells/well are utilized, except when samples are TNTC, in which case values generated from plating cells at 2.5×10^4 cells are used for that specific sample/stimuli/timepoint. Data processing as performed using the R programming language.

[00282] Intracellular cytokine assays are also performed. PBMCs are distributed at 1×10^6 cells per well into v-bottom 96-well plates. Cells are pelleted and resuspended in 100 μ l of complete RPMI containing the encoded peptides described above. DMSO is used as a negative control for each sample. Brefeldin A (Biolegend) is added to a final concentration of 5 μ g/mL after 1 hour and cells incubated overnight. Following viability stain, extracellular staining is performed in FACS buffer (PBS + 2% FBS + 2mM EDTA). Cells are washed, fixed and permeabilized with the eBiosciences Fixation/Permeabilization Solution Kit. Intracellular staining is performed. Samples are assessed for viability, CD3, CD4, CD8, IFN γ , TNF α , IL-2, Perforin, CD107a, CCR7, and CD45RA.

[00283] Serum cytokine markers are also monitored. Serum cytokine and chemokine levels are measured with standard multiplex assays. Serum is collected and marker analysis is performed at 0 hours (baseline), 2 hours, 8 hours, 24 hours and 48 hours following vaccination. Cytokines assessed are Interleukin-1 beta (IL-1 β), Interleukin-1 (IL-10), Interleukin-6 (IL-6), Tumor necrosis factor alpha (TNF- α), Interferon gamma (IFN- γ), Granulocyte-macrophage colony-stimulating factor (GM-CSF), Interferon gamma-induced protein 10 (IP-10), Monocyte chemoattractant protein-1 (MCP-1), Macrophage inflammatory protein 1 beta (MIP-1 β), and IFN-alpha (IFN- α 2a).

I. Immunogenicity Results in NHPs

[00284] Self-amplifying expression systems that include ssRNA vectors having modified nucleosides and/or having been purified using a chromatography system and/or an affinity-based separation system demonstrate improved vaccine efficacy as assessed by monitoring T cell and/or B cell responses.

Sequences

[00285] Vectors, cassettes, and antibodies referred to herein are described below and referred to by SEQ ID NO. Reference to SEQ ID NOS. is also with reference to the Sequence Listing found in U.S. Provisional Application No. 63/277,116, which is hereby incorporated by reference for all purposes.

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)
UbA76-25merPDTT nucleotide (SEQ ID NO:32)
UbA76-25merPDTT polypeptide (SEQ ID NO:33)
MAG-25merPDTT nucleotide (SEQ ID NO:34)
MAG-25merPDTT polypeptide (SEQ ID NO:35)
Ub7625merPDTT_NoSFL nucleotide (SEQ ID NO:36)
Ub7625merPDTT_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 [VEEV] (SEQ ID NO:3) GenBank: L01442.2
VEE-MAG25mer (SEQ ID NO:4); contains MAG-25merPDTT nucleotide (bases 30-1755)
Venezuelan equine encephalitis virus strain TC-83 [TC-83](SEQ ID NO:5) GenBank: L01443.1
VEEV Delivery Vector (SEQ ID NO:6); VEEV genome with nucleotides 7544-11175 deleted [alphavirus structural proteins removed, except the last 50 amino acids of E1]
ATGggcgcgcatgagagaagcccagaccaattacctacccaaaATGGagaaagttcacgttgacatcgaggaagacagcccattcctcag agctttgcagcggagcttcccgcagtttgaggtagaagccaagcaggtcactgataatgacatgctaatgccagagcgttttcgcatctggcttcaaaactgatcgaacggaggtggaccatccgacacgatccttgacattggaagtgcgccccccgcagaatgtattctaagcacaagtatcatt gtatctgtccgatgagatgtgcggaagatccggacagattgtataagatgcaactaagctgaagaaaaactgtaaggaaataactgataagga attggacaagaaaaatgaaggagctcgcgctcatgagcgcacctgacctggaaactgagactatgtgcctccacgacgacgagtcgtgtcgc tacgaagggcaagtcgctgtttaccaggatgtatacgcggtgacggaccgacaagtctctaccagcaataaggaggttagagtcgcta ctggataggcttgacaccacctttatgtttaagaactggctggagcatatccatcatactctaccaactgggcccgcgaaaccgtgtaacg gctcgtatacataggcctatgcagctctgacgttatggagcggtcacgtagaggatgtccattcttagaagaagtattgaaaccatccaacaat gttctattctctgttgctcgcacctctaccacgagaagagggacttactgaggagctggcacctgccgtctgtatttctacttacttggaagcaaa attacacatgctcggtgtgagactatagttagttgcgacgggtacgtcgttaaaagaatagctatcagtcaggcctgtatgggaagccttcaggct atgctgctacgatgcaccgcgagggtctgtgctgcaaaagtacagacacattgaacggggagagggtctctttccggtgtgcagtatgtgc cagctacattgtgtgaccaaataactgactggcactggcaacagatgtcagtgcgacgacgcgcaaaaactgctggtggctcaaccagcgtata gtcgtcaacggtcgcaccagagaaacaccaataacatgaaaaattacctttgccgtagtgcccaggcattgtctaggtgggcaaaaggaata taaggaagatcaagaagatgaaaggcactaggactacgagatagacagttagtcaggggtgtgtgggcttttagaaggcacaagataaca tctatttataagcggcgatcccaaaccatcatcaaagtgaacagcagatttccactcattcgtgctgccaggataggcagtaacacattggag atcgggctgagaacaagaatcaggaaaatgttagaggagcacaaggagccgtcacctctcattaccgaggagcgtacaagaagctaagtcgc gcagccgatgaggtaaggaggtgcgtgaagccgaggagttgcgagcagctctaccaccttggcagctgatgttggaggagcccactctggaag

ccgatgtcacttgatgttacaagaggctggggccggctcagtgagacacctcgtggcttgataaaggttaccagctacgctggcgaggacaag
atcggctcttacgctgtgctttctccgaggctgtactcaagagtgaataatcttgcacccctctcgtgaacaagtcagtgtataaac
actctggccgaaaaggcgttatgccgtggaaccataccatggtaaagtagtggtgccagaggacatgcaataccctccaggactttcaagct
ctgagtgaagtgccaccattgtgtacaacgaacgtgagttcgtaaacaggtacctgacacatattgccacacatggaggagcgtgaacactga
tgaagaatattcaaaaactgtcaagcccagcagcagcggcgaataacctgtacgacatcgacaggaacagtcgctcaagaaagaactagt
cactgggctagggtcacaggcagctggatcctccttccatgaattcgcctacgagagctgagaacacgaccagccgctccttaccagct
accaacataggggtgatggcgtgcaggatcaggcaagtctggcatcattaaaagcgcagtcacaaaaaagatctagtggtagcgcaa
gaaagaaaactgtgcagaaattataaggagctcaagaaaatgaaagggctggacgtcaatgccagaacttgactcagtgcttgaatgg
atgcaaacacccccctagagacctgtatattgacgaagctttgcttgcagcaggtactctcagagcgtcatagccattataagacctaaaa
ggcagtgctctcggggatcccaaacagtcgggttttttaacatgatgtcctgaaagtgattttaaccagagatttgacacaagcttccac
aaaagcatctctcgccgtgactaaatctgtgacttcggtcgtctcaacctgttttaacgacaaaaaatgagaacgacgaatccgaaagagact
aagattgtgattgactaccggcagtaacaaactaagcaggacgatctcattctcactgtttcagagggtgggtgaagcagttgcaaatagat
tacaagggcaacgaaataatgacggcagctgcctcctcaagggctgaccgtaaggtgtgatgccgttcggtacaaggtgaatgaaatcctct
gtacgacccacctcagaacatgtgaacgtcctactgacccgacggaggaccgcatcgtgtgaaaactagccggcgaccatggataaaa
aactgactgccaagtaccctgggaatttactgcccagatagaggagtggaagcagagcatgatgccatcatgaggcacatctggagagac
cggaccctaccgactctccagaataaggcaaacgtgtgtgggccaaggcttagtgccggtgctgaagaccgctggcatagacatgaccact
gaacaatggaacactgtggatttttgaacggacaagctcactcagcagagatagattgaaccaactatgctgagggtctttggactcgat
ctggactccggtctattttctgacccactgttccgttatcattaggaataatcactgggataactccccctcctaactgtacgggctgaataa
agaagtggtccgtcagctctctcgaggtaccacaactgcctcgggagcttgccaagagagctatgacatgaacactggtactgctgca
attatgatccgcgataaacctagctcctgtaaacagaagactgcctcatgcttagtctccaccataatgaacaccaagagtgacttttctc
attcgtcagcaaatgaaagggcagaactgtcctgggtgctggggaaaagtgtccgtcccaggcaaatggttgactggtgtcagaccggcctg
aggctaccttcagagctcggctggatttaggcatcccagggtgatgtgccaaatatgacataatattgttaatgtgaggacccatataaatacc
atcactatcagcagtgtaagaccatgccattaagcttagcatgttgaccaagaaagcttctgcatctgaatcccggcgaacctgtgtcagca
taggttatggttacgctgacagggccagcgaagcatcattggtgctatagcgggcagttcaagtttcccgggtatgcaaacgaaatcctcac
tgaagagacggaagtctgtttgattcattgggtacgatcgaaggccgtacgcacaatccttacaagcttcatcaaccttgaccaacatta
tacaggttcagactccaagcagcggatgtgacacctcatatcatgtggtgaggggatattgccacggcaccgaaggagtgattataatg
ctgtaacagcaaaaggacaacctggcggagggtgtgaggagcgtgtataagaaatcccggaaagcttctgattacagccgatcgaagtagg
aaaagcgcgactggtcaaaggtgagctaaacatatcattcatgctgtaggaccaaactcaacaaagttcggaggtgaaagtgacaaacag
ttggcagaggcttatgagtcacgtaagattgtcaacgataacaattacaagtcagtagcagattcactgtgtccaccggcatctttccggga
acaaagatcgactaaccaatcattgaaccattgctgacagcttagacaccactgatgcagatgtagccatatactgcagggacaagaaatgg
gaaatgactctcaaggaagcagtggttaggagagaagcagtgaggagatagcatatccgacgactctcagtgacagaacctgatgcagag
ctggtgagggtgatccgaagagttcttggctggaaggaaggctacagcacaagcagtggaacaaactttctcatattggaagggaccaagtt
tcaccaggcggccaaggatagcagaaattaatgcatgtggccggtgcaacggaggccaatgagcaggtatgcatgtatactcggagaa
agcatgagcagatttaggtcgaatgcccgtcgaagagtcggaagcctccacaccactagcacgctgcttctgtgcatcctgcatgact
ccagaaagagtacagcctaaaagcctcacgtccagaacaaattactgtgtcctatcttccattgccaagatagaaactggtgtgag
aagatccaatgctcccagcctatattgtctaccgaaagtcctgctgatatcctcaaggaagatctcgtggaacaccaccgtagacgag
actccggagccatcgagcagaaccaatccacagaggggacacctgaacaaccaccactataaccgaggatgagaccaggactagaacgct
gagccgatcatcgaagaggaagaagaggatagcataagttgctgtcagatggcccaccaccagggtgctgcaagtgcaggcagacattc
acgggccccctctgtatctagctcatctggtccattcctcatgcatccgactttgatgtggaagttatccatacttgacacctggaggagct
agcgtgaccagcggggcaacgtcagccgagactaactcttactctgcaagagatggagtttctggcgcgaccggtgctgctcgaacagt
attcaggaacctccacatcccgtccgcgcaagaacaccgtcacttgacccagcagggcctgctcgagaaccagcctagtttccacccgc
caggcgtgaatagggtgatcactagagaggagctcagggcgttaccctcagcactcctagcaggtcggctcagagaaccagcctggtctcc
aaccgccaaggcgtaaatagggtgattacaagagaggagttgaggcgttcgtagcacaacaacatgacggtttgatgagggtgacatactt
ttctccgacaccggtcaagggcatttacaacaaaaatcagtaaggcaaacggtgctatccgaagtggtgttgagaggaccgaattggagattt
cgtatgccccgcctcgaccaagaaaaagaagaattactacgcaagaattacagttaaatcccacacctgtaacagaagcagataaccagtc
caggaaggtggagaacatgaaagccataacagctagacgtattctgcaaggcctagggcattatttgaaggcagaaggaaaagtggagtgtca
ccgaacctgcatcctgttcttattcatctagtgtgaaccgtgcctttcaagcccaaggtcgcagtggaagcctgtaacgcatgttgaag
agaactttcgaactgtgcttctactgtattatccagagtagatgcctatttgacatggttgacggagcttcatgctgcttagacactgccagtt
ttgcccgtcaagagctgcgagctttcaaagaaacactcctatttggaaaccacaatacagatcggcagtgcttgcagatccagaacagctcc
agaactctggcagctgccaacaaaagaaattgcaatgtcacgcaaatgagagaattgccctattggattcggcggcctttaatgtggaatgc
tcaagaaatgctgttaataatgaatattgggaaacgtttaaagaaaacccatcaggcttactgaagaaaactggttaaatcattaccaa
attaaaaggacaaaagctgctgcttttgcgaagacacataattgaaatgttgacagacataccaatggacaggttgaatggacttaa
gagagagctgaaagtgactccaggaacaaaacatactgaagaacggccaaggtacaggtgatccaggctccgatccgtagcaacagcgt

GGAGGACCGCATCGTGTGGAAAACACTAGCCGGCGACCCATGGATAAAAACACTGACTGCCAAGT
ACCCTGGGAATTTCACTGCCACGATAGAGGAGTGGCAAGCAGAGCATGATGCCATCATGAGGCAC
ATCTTGAGAGACCGGACCCTACCGACGTCTTCCAGAATAAGGCAAACGTGTGTTGGGCCAAGGCT
TTAGTGCCGGTGCTGAAGACCGCTGGCATAGACATGACCACTGAACAATGGAACACTGTGGATTAT
TTTGAACCGGACAAAGCTCACTCAGCAGAGATAGTATTGAACCAACTATGCGTGAGGTTCTTTGGA
CTCGATCTGGACTCCGGTCTATTTTCTGCACCCACTGTTCCGTTATCCATTAGGAATAATCACTGGG
ATAACTCCCCGTCGCCTAACATGTACGGGCTGAATAAAGAAGTGGTCCGTCAGCTCTCTCGCAGGT
ACCCACAACCTGCCTCGGGCAGTTGCCACTGGAAGAGTCTATGACATGAACACTGGTACACTGCGCA
ATTATGATCCGCGCATAAACCTAGTACCTGTAAACAGAAGACTGCCTCATGCTTTAGTCCTCCACCA
TAATGAACACCCACAGAGTGACTTTTTCTTCATTCGTGAGCAAAATTGAAGGGCAGAAGCTGCTGCT
GGTCGGGGAAAAGTTGTCCGTCAGGCAAAATGGTTGACTGGTTGTGACACCGGCTGAGGCTAC
CTTCAGAGCTCGGCTGGATTTAGGCATCCCAGGTGATGTGCCCAAATATGACATAAATTTGTTAAT
GTGAGGACCCCATATAAATACCATCACTATCAGCAGTGTGAAGACCATGCCATTAAGCTTAGCATG
TTGACCAAGAAAAGCTTGTCTGCATCTGAATCCCGGGCGAACCTGTGTCAGCATAGGTTATGGTTAC
GCTGACAGGGCCAGCGAAAGCATCATTGGTGTATAGCGCGGCAGTTCAAGTTTTCCCGGGTATGC
AAACCGAAATCCTCACTTGAAGAGACGGAAGTTCTGTTTGTATTCAATTGGGTACGATCGCAAGGCC
CGTACGCACAATCCTTACAAGCTTTCATCAACCTTGACCAACATTTATACAGGTTCCAGACTCCACG
AAGCCGGATGTGCACCCTCATATCATGTGGTGCAGGGGATATTGCCACGGCCACCGAAGGAGTG
ATTATAAATGCTGCTAACAGCAAAGGACAACCTGGCGGAGGGGTGTGCGGAGCGCTGTATAAGAA
ATTCCCGGAAAGCTTCGATTTACAGCCGATCGAAGTAGGAAAAGCGCGACTGGTCAAAGGTGCAG
CTAAACATATCATTATGCCGTAGGACCAAACTTCAACAAAGTTTCGGAGGTTGAAGGTGACAAAC
AGTTGGCAGAGGCTTATGAGTCCATCGCTAAGATTGTCAACGATAACAATTACAAGTCAGTAGCGA
TTCCACTGTTGTCCACCGGCATCTTTCCGGGAACAAAGATCGACTAACCCAATCATTGAACCATTT
GCTGACAGCTTTAGACACCACTGATGCAGATGTAGCCATATACTGCAGGGACAAGAAATGGGAAA
TGACTCTCAAGGAAGCAGTGGCTAGGAGAGAAGCAGTGGAGGAGATATGCATATCCGACGACTCT
TCAGTGACAGAACCTGATGCAGAGCTGGTGAGGGTGCATCCGAAGAGTTCTTTGGCTGGAAGGAA
GGGCTACAGCACAAGCGATGGCAAACTTTCTCATATTTGGAAGGGACCAAGTTTCACCAGGCGGC
CAAGGATATAGCAGAAATTAATGCCATGTGGCCCGTTGCAACGGAGGCCAATGAGCAGGTATGCA
TGTATATCCTCGGAGAAAGCATGAGCAGTATTAGGTCGAAATGCCCCGTCGAAGAGTCCGGAAGCCT
CCACACCACCTAGCACGCTGCCTTGCTTGTGCATCCATGCCATGACTCCAGAAAAGGTACAGCGCC
TAAAAGCCTCACGTCCAGAACAAATTACTGTGTGCTCATCCTTTCCATTGCCGAAGTATAGAATCAC
TGGTGTGCAGAAGATCCAATGCTCCCAGCCTATATTGTTCTCACCGAAAAGTGCCTGCGTATATTCAT
CCAAGGAAGTATCTCGTGGAACACCACCGGTAGACGAGACTCCGGAGCCATCGGCAGAGAACCA
ATCCACAGAGGGGACACCTGAACAACCACCAATTATAACCGAGGATGAGACCAGGACTAGAACGC
CTGAGCCGATCATCATCGAAGAGGGAAGAAGAGGATAGCATAAAGTTTGTGTGCTGAGTGGCCCGACC
CACCAGGTGCTGCAAGTCGAGGCAGACATTCACGGGCGCCCTCTGTATCTAGCTCATCCTGGTCC
ATTCTCATGCATCCGACTTTGATGTGGACAGTTTATCCATACTTGACACCCTGGAGGGAGCTAGCG
TGACCAGCGGGGCAACGTCAGCCGAGACTAACTCTTACTTTCGCAAAGAGTATGGAGTTTCTGGCGC
GACCGGTGCCTGCGCCTCGAACAGTATTAGGAACCCTCCACATCCCGCTCCGCGCACAAAGAACAC
CGTCACTTGCACCCAGCAGGGCCTGCTCGAGAACCAGCCTAGTTTCCACCCCGCCAGGCCTGAATA
GGGTGATCACTAGAGAGGAGCTCGAGGCGCTTACCCCGTCACGCACTCCTAGCAGGTCCGTTCTCGA
GAACCAGCCTGGTCTCCAACCCGCCAGGCGTAAATAGGGTGATTACAAGAGAGGAGTTTGGGGCG
TTCGTAGCACACAACAATGACGTTTTGATGCGGGTGCATACATCTTTTCTCCGACACCGGTCAA
GGGCATTTACAACAAAATCAGTAAGGCAAACGGTGTATCCGAAGTGGTGTGGAGAGGACCGA
ATTGGAGATTTCTGATGCCCCGCGCCTCGACCAAGAAAAAGAAGAATTACTACGCAAGAAATTACA
GTTAAATCCCACACCTGCTAACAGAAGCAGATACCAGTCCAGGAAGGTGGAGAACATGAAAGCCA
TAACAGCTAGACGTATTCTGCAAGGCCTAGGGCATTATTTGAAGGCAGAAGGAAAAGTGGAGTGC
TACCGAACCTGCATCCTGTTCTTTGTATTATCTAGTGTGAACCGTGCCTTTTCAAGCCCAAGG
TCGCAGTGGAAGCCTGTAACGCCATGTTGAAAGAGAACTTTCCGACTGTGGCTTCTTACTGTATTAT
TCCAGAGTACGATGCCTATTTGGACATGGTTGACGGAGCTTCATGCTGCTTAGACACTGCCAGTTTT
TGCCCTGCAAAGCTGCGCAGCTTTCCAAAGAAACACTCCTATTTGGAACCCACAATACGATCGGCA
GTGCCTTACGCGATCCAGAACACGCTCCAGAACGTCCTGGCAGCTGCCACAAAAAGAAATTGCAAT
GTCACGCAAAATGAGAGAATTGCCCGTATTGGATTCCGGCGCCTTTAATGTGGAATGCTTCAAGAAA
TATGCGTGAATAATGAATATTGGGAAACGTTTAAAGAAAACCCCATCAGGCTTACTGAAGAAAAC
GTGGTAAATTACATTACCAAATTTAAAGGACCAAAAAGCTGCTGCTCTTTTTGCGAAGACACATAAT
TTGAATATGTTGCAGGACATAACCAATGGACAGGTTTGTAAATGGACTTAAAGAGAGACGTGAAAGTG
ACTCCAGGAACAAAACATACTGAAGAACGGCCAAAGGTACAGGTGATCCAGGCTGCCGATCCGCT
AGCAACAGCGTATCTGTGCGGAATCCACCGAGAGCTGGTTAGGAGATTAATGCGGCTCCTGCTTCC
GAACATTACATACACTTTGATATGTCGGCTGAAGACTTTGACGCTATTATAGCCGAGCATTCCAG
CCTGGGGATTGTGTTCTGGAACTGCATCGCGTCGTTTGATAAAAAGTGAGGACGACGCCATGGCT
CTGACCCGCTTAATGATTCTGGAAGACTTAGGTGTGGACGCAGAGCTGTTGACGCTGATTGAGGCG
GCTTTCGGCGAAATTTTCATCAATACATTTGCCCACTAAAACATAAATTTAAATTCGGAGCCATGATGA

AATCTGGAATGTTCTCACACTGTTTGTGAACACAGTCATTAACATTGTAATCGCAAGCAGAGTGTT
GAGAGAACGGCTAACCGGATACCATGTGCAGCATTGATTGGAGATGACAATATCGTGAAAGGAG
TCAAATCGGACAAATTAATGGCAGACAGGTGCGCCACCTGGTTGAATATGGAAGTCAAGATTATAG
ATGCTGTGGTGGGCGAGAAAGCGCCTTATTTCTGTGGAGGGTTTATTTTGTGTGACTCCGTGACCGG
CACAGCGTGCCGTGTGGCAGACCCCTAAAAAGGCTGTTAAGCTTGGCAAACCTCTGGCAGCAGA
CGATGAACATGATGATGACAGGAGAAGGGCATTGCATGAAGAGTCAACACGCTGGAACCGAGTGG
GTATTCTTTCAGAGCTGTGCAAGGCAGTAGAATCAAGGTATGAAACCGTAGGAACTTCCATCATAG
TTATGGCCATGACTACTCTAGCTAGCAGTGTTAAATCATTACAGCTACCTGAGAGGGGCCCTATAA
CTCTCTACGGCTAACCTGAATGGACTACGACATAGTCTAGTCCGCCAAGATGTTCCCGTTCAGCCA
ATGTATCCGATGCAGCCAATGCCCTATCGCAACCCGTTCCGCGCCCGCGCAGGCCCTGGTTCCTCC
AGAACCGACCTTTTCTGGCGATGCAGGTGCAGGAATTAACCCGCTCGATGGCTAACCTGACGTT
AAGCAACGCGGGACGCGCCACCTGAGGGGCCATCCGCTAAGAAACCGAAGAAGGAGGCCCTCGCA
AAAACAGAAAGGGGGAGGCCAAGGGAAGAAGAAGAACAAGGGAAGAAGAAGGCTAAGAC
AGGGCCGCTAATCCGAAGGCACAGAATGGAAACAAGAAGAAGACCAACAAGAAACCAGGCAAG
AGACAGCGCATGGTCATGAAATTGGAATCTGACAAGACGTTCCCAATCATGTTGGAAGGGAAGAT
AAACGGCTACGCTTGTGTGGTTCGGAGGGAAGTTATTCAGGCCGATGCATGTGGAAGGCAAGATCG
ACAACGACGTTCTGGCCGCGCTTAAGACGAAGAAAGCATCCAAATACGATCTTGAGTATGCAGATG
TGCCACAGAACATGCGGGCCGATACATTCAAATACACCCATGAGAAACCCCAAGGCTATTACAGCT
GGCATCATGGAGCAGTCCAATATGAAAATGGGCGTTTCACGGTGCCGAAAGGAGTTGGGGCCAAG
GGAGACAGCGGACGACCCATTCTGGATAACCAGGGACGGGTGGTTCGCTATTGTGCTGGGAGGTGT
GAATGAAGGATCTAGGACAGCCCTTTCAGTCGTATGTGGAACGAGAAGGGAGTTACCGTGAAGT
ATACTCCGGAGAACTGCGAGCAATGGTCACTAGTGACCACCATGTGTCTGCTCGCCAATGTGACGT
TCCCATGTGCTCAACCACCAATTTGCTACGACAGAAAACCAGCAGAGACTTTGGCCATGCTCAGCG
TTAACGTTGACAACCCGGGCTACGATGAGCTGCTGGAAGCAGCTGTTAAGTGCCCCGGAAGGAAA
AGGAGATCCACCGAGGAGCTGTTAATGAGTATAAGCTAACCGGCCCTTACATGGCCAGATGCATC
AGATGTGCAGTTGGGAGCTGCCATAGTCCAATAGCAATCGAGGCAGTAAAGAGCGACGGGCACGA
CGGTTATGTTAGACTTCAGACTTCCTCGCAGTATGGCCTGGATTCCCTCCGGCAACTTAAAGGGCAG
GACCATGCGGTATGACATGCACGGGACCATTAAGAGATAACCACTACATCAAGTGCTACTCTATAC
ATCTCGCCCGTGTACATTGTGGATGGGCACGGTTATTTCTGCTTGCCAGGTGCCCGGCAGGGGA
CTCCATCACCATGGAATTAAGAAAGATTCCGTCAGACACTCCTGCTCGGTGCCGTATGAAGTGAA
ATTTAATCCTGTAGGCAGAGAACTCTATACTCATCCCCAGAACACGGAGTAGAGCAAGCGTGCCA
AGTCTACGCACATGATGCACAGAACAGAGGAGCTTATGTGAGATGCACCTCCCGGGCTCAGAAGT
GGACAGCAGTTTGGTTTCTTGAGCGGCAGTTCAGTCACCGTGACACCTCCTGATGGGACTAGCGC
CCTGGTGAATGCGAGTGTGGCGGCACAAAGATCTCCGAGACCATCAACAAGACAAAACAGTTCA
GCCAGTGACAAAGAAGGAGCAGTGCAGAGCATATCGGCTGCAGAACGATAAGTGGGTGTATAAT
TCTGACAAACTGCCAAAGCAGCGGGAGCCACCTTAAAGGAAAACCTGCATGTCCCATTCTTGCTG
GCAGACGGCAAATGCACCGTGCCTCTAGCACAGAACCTATGATAACCTTCGGTTTCAGATCAGTG
TCACTGAAACTGCACCTAAGAATCCCACATATCTAATCACCCGCCAACTTGCTGATGAGCCTCACT
ACACGCACGAGCTCATATCTGAACCAGCTGTTAGGAATTTTACCGTCACCGAAAAAGGGTGGGAGT
TTGTATGGGGAAACCACCCGCCGAAAAGTTTTGGGCACAGGAAACAGCACCCGGAAATCCACAT
GGGCTACCGCACGAGGTGATAACTCATTATTACCACAGATACCCTATGTCCACCATCCTGGGTTTGT
CAATTTGTGCCGCCATTGCAACCGTTTCCGTTGCAGCGTCTACCTGGCTGTTTTGCAGATCTAGAGT
TGCGTGCCTAACTCCTTACCGGCTAACACCTAACGCTAGGATAACATTTTGTCTGGCTGTGCTTTC
TGCGCCCGCACTGCCCGGGCCGAGACCACCTGGGAGTCTTGGATCACCTATGGAACAATAACCA
CAGATGTTCTGGATTCAATTGCTGATCCCTCTGGCCGCTTGTATCGTAGTACTCGCCTGCTCAGGT
GCGTGTGCTGTGCTGCTGCTTTTTTAGTCATGGCCGCGCCGAGGCGCCGGCGCCTACGAGCACG
CGACCACGATGCCGAGCCAAGCGGGAATCTCGTATAACACTATAGTCAACAGAGCAGGCTACGCA
CCACTCCCTATCAGCATAACACCAACAAGATCAAGCTGATACCTACAGTGAACCTGGAGTACGTC
ACCTGCCACTACAAAACAGGAATGGATTACCAGCCATCAAATGCTGCGGATCTCAGGAATGCACT
CCAACCTACAGGCCTGATGAACAGTGCAAAGTCTTACAGGGGTTTACCCGTTTCATGTGGGGTGGT
GCATATTGCTTTTGCACACTGAGAACACCCAAGTCAGCAAGGCCTACGTAATGAAATCTGACGAC
TGCTTGCAGGATCATGCTGAAGCATATAAAGCGCACACAGCCTCAGTGCAGGCGTTCCCTCAACATC
ACAGTGGGAGAACACTCTATTGTGACTACCGTGTATGTGAATGGAGAACTCCTGTGAATTTCAAT
GGGGTCAAAATAACTGCAGGTCCGCTTTCCACAGCTTGGACACCCCTTGTATCGAAAATCGTGCAG
TATGCCGGGGAGATCTATAATTATGATTTTCTGAGTATGGGGCAGGACAACCAGGAGCATTTGGA
GATATACAATCCAGAACAGTCTCAAGCTCTGATCTGTATGCCAATACCAACCTAGTGTGTCAGAGA
CCCAAAGCAGGAGCGATCCACGTGCCATACACTACCGCACCTTCGGGTTTTGAGCAATGGAAGAA
AGATAAAGTCCATCATTGAAATTTACCGCCCTTTCGGATGCGAAATATATACAACACCCCTTCG
CGCCAAAACCTGTGCTGATGGGTCAATTTCCATTAGCCTTTGACATTTCCGACGCTTTGTTCCACAGG
GTGTCAGAAACACCGCACACTTTACGCGCCGAATGCACACTCTTAACGAGTGCCTGTATTCTCCGAC
TTTGGTGGGATCGCCACGGTCAAGTACTCGGCCAGCAAGTCAGGCAAGTGCAGTCCATGTGCCA
TCAGGGACTGCTACCTAAAAGAAGCAGCAGTCGAGCTAACCGAGCAAGGGTTCGGCGACTATCCA

TTTCTCGACCCGCAAATATCCACCCGGAGTTTCAGGCTCCAAATATGCACATCATATGTTACGTGCAA
AGGTGATTGTCACCCCCGAAAAGACCATATTGTGACACACCCTCAGTATCACGCCAAACATTTAC
AGCCGCGGTGTCAAAAACCGCGTGGACGTGGTTAACATCCCTGCTGGGAGGATCAGCCGTAATTAT
TATAATTGGCTTGGTGTGGCTACTATTGTGGCCATGTACGTGCTGACCAACCAGAAAACATAATTG
AATACAGCAGCAATTGGCAAGCTGCTTACATAGAACTCGCGGCGATTGGCATGCCGCCTTAAATTT
TTTATTTTATTTTTCTTTTCTTTTCCGAATCGGATTTTGTFTTTAATATTTTC

VEEV Production Vector (SEQ ID NO:8); VEEV genome with nucleotides 7544-11175 deleted, plus 5' T7-
promoter, plus 3' restriction sites

TAATACGACTCACTATAGGATGggcggcgcatgagagaagcccagaccaattacctacccaaaATGGagaaagttcacgttgacatcga
ggaagacagcccattcctcagagctttgcagcggagcttcccgaatttgaggtagaagccaagcaggtcactgataatgacctgctaagccagagcgttttgc
atctgcttcaaaactgatcgaacggaggtggacccatccgacacgatccttgacattggaagtgcgccgccgagaatgatttcaagcacaagatcatgta
tctgtccgatgagatgtcggaaalccggacagattgataagtgaactaagctgaagaaaaactgtaaggaaataactgataagaaatggacaagaaaatg
aaggagctcggccgcatgagcagacctgacctggaaactgagactatgtccctcacgacgacgagtcgtgtcgtacgaagggcaagtgcgtgttaccag
gatgtatacgcggtgacggaccgacaagtctatccaagccaataaggagttagatgcctactgataggctttgacaccacctttatgtttaagaactt
ggctggagcatalccatcactctaccaactggcggcagaaaccgtgtaacggctgtaacataggcctatgcagctctgacctgtagaggagctggcactg
ggatgtccattctagaagaagtattgaaacctcaacaatgttcttctgttggctcaccatcaccacgagaagagactgactgaggagctggcactg
ccgtctgatttccattcgtggcaagcaaaattacacatgtcgtgtgagactatagttgctgacgggtacgtcgttaaagaatagctatcagtcaccagctg
tgggaagcctttagctatgctgctacatgcaccgcgagggtattctgtgctgcaaaagtgcagacacattgaacgggagaggctcttctccgtgtcacgt
atgtgccagctacattgtgtgaccaaatgactggcactaggcaacagatgtcagtgccgacgacgcgcaaaactgctggtggctcaaccagctatagtcgtc
aacggtcgcaccagagaacaccaatccatgaaaaattacctttgcccgtagtgcccaggcatttgcagtgaggcaaaagaaataaggaagatcaagaag
atgaaaggccactagactacgagatagacagttagtcagttgggtgtgtggcttttagaaggcacaagataacatctattataagcggccgatacccaacca
tcatcaaaagtgaacagcgaattccactcattctgctgcccaggataggcagtaacacattggagatcgggtgagaacaagaatcaggaaaatgttagaggca
caaggagccgtcacctctcattaccggcaggactacaagaagtaagtgcgacgccgatgaggctaaggaggtgctgaaagccgaggagttgcgagcagct
ctaccacctttggcagctgatgttaggagcccactctggaagccgatgtcactgatgttacaagaggctggggccggctcagtgagacacctctggtgat
aaagttaccagctacgctggcaggacaagatcggctctacgctgtgcttctccgagcgtgtactcaagagtgaaaaattatctgcatccacctctcgtgaa
caagtcatagtgataacactctggccaaaaggcgttatgcccgtggaaccataccatgtaaaagtagtggtccagagggacatgcaataccgctccaggact
tcaagctctgagtgaagtgccaccattgttacaacgaactgagttcgtaaacaggtacctgacatattgccacataggaggagcgtgaactgatgaa
gaatattcaaaaactgtcaagcccagcgagcagcagcgcaalacctgacacatgcagaaagaaactgctcaagaagaactgactgggctagggtc
cacaggcgagctgtgtgatcctccctccatgaattcgcctacgagagctgagaacacgaccagccgctccttaccagtaaccaatagggtgtatggcgtg
ccagatcaggaagctgtgcatcattaaaagcgcagtcaccaaaaagatctagtgtgagcgccaagaaaagaaactgtgcagaattataaggagcgtcaag
aaaatgaaaggctgtgagctcaatgccagaactgtgactcagtccttgaatggatgcaaacaccccgtagagacctgtatattgacgaagcttttgcctgtcat
gcaggtactctagagcgtcatagccattataagacctaaaaggcagtgctctgcccggatcccaaacagtgccgtttttaacatgatgtcctgaaagtgact
ttaaccacgagatttgcacacaagcttcccaaaaagcactctcgcctgtgactaaactgtgactcggctgtcicaacctgttttacgacaaaaaatgagaacg
acgaatccgaaagagactaagattgtattgacactaccggcagtaaccaaacctaaagcaggacgatctattcactgtttcagaggggtgggtgaagcattgca
aatagattacaaggcaacgaaataatgacggcagctgctcctcaagggtgacctgaaaggtgtgtatgccgttccggtacaaaggtgaatgaaaatcctctgacg
caccacctcagaactgtgaactcactgaccgcagggagaccgcatcgtgtgaaaacactagccggcaccatggataaaaactgactgccaag
tacctgggaatttactgcccagatagaggatggcaagcagagcatgatgccatcagaggcactctggagagaccggacctaccgactcttccagaata
aggcaaacgtgtgtggccaaggctttatgcccgtgctgaagaccgctgcatagacatgacctgaaactggaactgtgattatttgaaacggcaaaa
ctcactcagcagatagattgaaccaactatgctgaggtcttggactgactgactcctcgcaggtaccacaactgctcggcctgactgccaactgccaag
agtctatgacatgaacactgtacactgcgaattatgatccgcataaaactagctgtaaacagaagactgctcactgcttgaatgctcctcaccataatgaacac
ccacagagtgactttctcattctcagcaaatgaaaggcagaactgctcctggtgtcgggaaaagttgtccgtccagcaaaatggttactggtgtcagac
cgccctgaggctacctcagagctcggctgatttaggcatcccagggtgattgcccataatgacataatattgttaattgtaggaccatataataccatcacta
tcagcagtgtagagccatgccattaaagcttagcatgtgaccaagaaagctgtctgactgaaatcccggcgaacctgtgtagcataggttatgttactgctgac
aggccagcgaagcattggtgtatagcggcagttcaagtttcccgggtatgcaaacccgaaatcctcactgaaagagcggaggttctgttattcattg
ggtacatgcgaagcccgtacgacaactcttacaagcttcatcaacttgaccaacatttatacaggttccagactccagaaagccggatgtcacctcatalc
atgtgtgaggggatattgccagggcaccgaaggagtattataaactgctgtaacagcaaaaggacaacctggcggaggggtgtcggagcgtgtataag
aaatcccggaaagcttcgattacagccgatcgaagtggaaaagcgcactgtgcaaaagtgtagtaaacataatcattcatccgtaggacaaactcaaca
agtctcgaggttgaaggtgacaacagttggcagaggttatgattccatcgtaaagttgtaaacgataaacatcaagtcagtagcattcactgtgtccacc
ggcatctttccggaaacaagatcactaaccaatcattgaaccattgctgacagctttgacaccactgatgcagatgtagcatatactgacgggacaagaaa
tggaaatgactctcaaggaaagcagtgctaggagagaagcagtgaggagatgcatatccgacgactctcagtgacagaacctgatgacagctgtgag
ggtgcatccgaagagttcttggctggaaggaaaggctacagcacaagcagtgcaaaacttctcatatttgaaggaccaggttccaccagcgcccaaggat
atagcagaataatgccatgtgcccgttgcacggagccaatgagcaggtatgcatgtatcctcggagaaagcagtagcagatttaggtcgaatgccccgt
cgaagagtcggaagcctccacaccactgacgctgcttgcctatccatgccatgactccagaaagtagtagcgcctaaaaagcctcacgtccagaaca
aattactgtgtcctatcttccattgccgaagtatagaatcactggtgtgcagaagatccaatgctcccagcctatattgttctaccgaaagtgcctgctatattcat
ccaaggaaatctctgtgaaacaccaccgtagacgagactccggagccatcggcagagaaccaatccacagaggggacacctgaacaaccaccactataaa
ccgagatgagaccaggactagaacgctgagccgatcatcagaagaggaaagaggatgataagttgctgtagatggcccagaccaccaggtgctg
caagtcgaggcagacatcagggccgcccctctgtatctagctatcctgtgctcattctctatgcatccgactttgatgtgacagtttaccacttgaaccctgga
gggagctagcgtgaccagcggggcaactgacggagactaactctactgcaaaagatgagtttctggcgcgaccggtgctgcgcctcgaacagtattc
aggaaacctccatcccgtccgcgcaagaacaccgtcactgaccaccagcagggcctgctcagaaccagcctgattccaccggcaggcgtgaatag
ggtgactactagagagagctcagggcgttaccctcagcactcctagcaggtcggctcagaaccagcctgctcacaaccggcaggcgttaaatagggt
gattacaagagaggagttgagggctctagcacaacaacaatgacggtttgatcgggtgcatacatcttctccgacaccggcgaagggcattacaacaaa

AAGCATCTCTCGCCGTTGCACTAAATCTGTGACTTCGGTTCGTCTCAACCTTGTTTTACGACAAAAA
 ATGAGAACGACGAATCCGAAAGAGACTAAGATTGTGATTGACACTACCGGCAGTACCAAACCTAA
 GCAGGACGATCTCATTCTCACTTGTTCAGAGGGTGGGTGAAGCAGTTGCAAATAGATTACAAAGG
 CAACGAAATAATGACGGCAGCTGCCTCTCAAGGGCTGACCCGTAAAGGTGTGTATGCCGTTCCGTA
 CAAGGTGAATGAAAATCCTCTGTACGCACCCACCTCAGAACATGTGAACGTCTACTGACCCGCAC
 GGAGGACCGCATCGTGTGAAAAACTAGCCGGCGACCCATGGATAAAAACTGACTGCCAAGT
 ACCCTGGGAATTTCACTGCCACGATAGAGGAGTGGCAAGCAGAGCATGATGCCATCATGAGGCAC
 ATCTTGAGAGACCGGACCCTACCGACGTCTTCCAGAATAAGGCAAACGTGTGTTGGGCCAAGGCT
 TTAGTGCCGGTGTGAAGACCGCTGGCATAGACATGACCATAAACAATGGAACACTGTGGATTAT
 TTTGAAACGGACAAAGCTCACTCAGCAGAGATAGTATTGAACCAACTATGCGTGAGGTTCTTTGGA
 CTCGATCTGGACTCCGGTCTATTTTCTGCACCCACTGTTCGGTTATCCATTAGGAATAATCACTGGG
 ATAACTCCCCGTGCTAACATGTACGGGCTGAATAAAGAAGTGGTCCGTGAGTCTCTCGCAGGT
 ACCACAACCTGCCTCGGGCAGTTGCCACTGGAAGAGTCTATGACATGAACACTGGTACACTGCGCA
 ATTATGATCCGCGCATAAACCTAGTACCTGTAAACAGAAGACTGCCTCATGCTTTAGTCCTCCACCA
 TAATGAACACCCACAGAGTGACTTTTTCTTCAATTCGTGACGAAATTGAAGGGCAGAACTGTCCTGGT
 GGTCGGGGAAAAGTTGTCCGTCCCAGGCAAATGGTTGACTGGTTGTCAGACCCGGCCTGAGGCTAC
 CTTTCAAGGCTCGGCTGGATTTAGGCATCCCAGGTGATGTGCCCAAATATGACATAATTTGTTAAT
 GTGAGGACCCCATATAAATACCATCACTATCAGCAGTGTGAAGACCATGCCATTAAGCTTAGCATG
 TTGACCAAGAAAGCTTGTCTGCATCTGAATCCCGGCGGAACCTGTGTGAGCATAGGTTATGGTTAC
 GCTGACAGGGCCAGCGAAAGCATCATTGGTGTATAGCGCGGCAGTTCAAGTTTTCCCGGGTATGC
 AAACCGAAATCCTCACTTGAAGAGACGGAAGTTCTGTTTGTATTCAATGGGTACGATCGCAAGGCC
 CGTACGCACAATCCTTACAAGCTTTCATCAACCTTGACCAACATTTATACAGGTTCCAGACTCCACG
 AAGCCGGATGTGCACCCTCATATCATGTGGTGCAGGGGATATTGCCACGGCCACCGAAGGAGTG
 ATTATAAATGCTGCTAACAGCAAAGGACAACCTGGCGGAGGGGTGTGCGGAGCGCTGTATAAGAA
 ATTCCCGGAAAGCTTCGATTTACAGCCGATCGAAGTAGGAAAAGCGCGACTGGTCAAAGGTGCAG
 CTAACATATCATTATGCCGTAGGACCAAACCTTCAACAAAGTTTTCCGAGGTTGAAGGTGACAAAC
 AGTTGGCAGAGGCTTATGAGTCCATCGCTAAGATTGTCAACGATAACAATTACAAGTCAGTAGCGA
 TTCCACTGTTGTCCACCCGCATCTTTTCCGGGAACAAAGATCGACTAACCCAATCATTGAACCATTT
 GCTGACAGCTTTAGACACCACTGATGCAGATGTAGCCATATACTGCAGGGACAAGAAATGGGAAA
 TGACTCTCAAGGAAGCAGTGGCTAGGAGAGAAGCAGTGGAGGAGATATGCATATCCGACGACTCT
 TCAGTGACAGAACCTGATGCAGAGCTGGTGAAGGTGCATCCGAAGAGTTCCTTTGGCTGGAAGGAA
 GGGCTACAGCACAAGCGATGGCAAACCTTCTCATATTTGGAAGGGACCAAGTTTACCAGGCGGC
 CAAGGATATAGCAGAAATTAATGCCATGTGGCCCGTTGCAACGGAGGCCAATGAGCAGGTATGCA
 TGTATCCTCGGAGAAAGCATGAGCAGTATTAGTTCGAAATGCCCGTCGAAGAGTTCGGAAGCCT
 CCATACCCTAGCAGCTGCCTTGCTGTGCTATGCCATGCCATGACTCCAGAAAGAGTACAGCGCC
 TAAAAGCCTCAGTCCAGAACAAATTAAGTGTGTGCTATCCTTTCCATTGCCGAAGTATAGAATCAC
 TGGTGTGAGAAAGATCCAATGCTCCAGCCTATATTGTTCTCACCGAAAGTGCCTGCGTATATTCAT
 CCAAGGAAGTATCTCGTGGAAACACCACCGGTAGACGAGACTCCGGAGCCATCGGCAGAGAACCA
 ATCCACAGAGGGGACACCTGAACAACCACCTTATAACCGAGGATGAGACCAGGACTAGAACGC
 CTGAGCCGATCATCATCGAAGAGGAAGAAGAGGATAGCATAAGTTTGTGCTGTCAGATGGCCCCGACC
 CACCAGGTGCTGCAAGTCGAGGCAGACATTCACGGGCGGCCCTCTGTATCTAGCTCATCCTGGTCC
 ATTCTCATGCATCCGACTTTGATGTGGACAGTTTATCCATACTTGACACCCTGGAGGGAGCTAGCG
 TGACCAGCGGGCAACGTCAGCCGAGACTAACTCTTACTTCCGAAAGAGTATGGAGTTTCTGGCGC
 GACCGGTGCCTGCGCCTCGAACAGTATTCAGGAACCCTCCACATCCCGCTCCGCGCACAGAACAC
 CGTCACTTGCACCCAGCAGGGCCTGCTCGAGAACCAGCCTAGTTTCCACCCCGCCAGGCGTGAATA
 GGGTGTACTAGAGAGGAGCTCGAGGCGCTTACCCCGTCACGCACTCCTAGCAGGTCCGTCTCGA
 GAACCAGCCTGGTCTCCAACCCGCGAGGCGTAAATAGGGTGATTACAAGAGAGGAGTTTGAAGCG
 TTCGTAGCACACAACAATGACGGTTTGTGCGGGTGCATACATCTTTTCTCCGACACCGGTCAA
 GGGCATTTACAACAAAATCAGTAAGGCAAACGGTGTATCCGAAGTGGTGTGGAGAGGACCGA
 ATTGGAGATTTGATGCCCCGCGCCTCGACCAAGAAAAAGAAATTACTACGCAAGAAATTACA
 GTTAAATCCCACACCTGCTAACAGAAGCAGATACCAGTCCAGGAAGGTGGAGAACATGAAAGCCA
 TAACAGCTAGACGATTCTGCAAGGCCTAGGGCATTATTTGAAGGCAGAAGGAAAAGTGGAGTGC
 TACCGAACCTGCATCCTGTTTCTTGTATTATCTAGTGTGAACCGTGCCTTTTCAAGCCCCAAGG
 TCGCAGTGAAGCCTGTAAACGCCATGTTGAAAGAGAAGTTTCCGACTGTGGCTTCTTACTGTATTAT
 TCCAGAGTACGATGCCTATTTGGACATGGTTGACGGAGCTTCATGCTGCTTAGACACTGCCAGTTTT
 TGCCCTGCAAAGCTGCGCAGCTTTCCAAAGAAACACTCCTATTTGGAACCCACAATACGATCGGCA
 GTGCCCTCAGCGATCCAGAACACGCTCCAGAAGTCCGCGAGCTGCCACAAAAGAAATTGCAAT
 GTCACGCAATGAGAGAATTGCCCGTATTGGATTCCGGCGCCTTTAATGTGGAATGCTTCAAGAAA
 TATGCGTGAATAATGAATATTGGGAAACGTTTAAAGAAAACCCCATCAGGCTTACTGAAAGAAAC
 GTGGTAAATTACATTACCAAAATTAAGGACCAAAAAGCTGCTGCTCTTTTTCGAAAGACACATAAT
 TTGAATATGTTGACAGGACATAACCAATGGACAGGTTTGTAAATGGACTTAAAGAGAGACGTGAAAGTG
 ACTCCAGGAACAAAACATACTGAAGAACGGCCCAAGGTACAGGTGATCCAGGCTGCCGATCCGCT

AGCAACAGCGTATCTGTGCGGAATCCACCGAGAGCTGGTTAGGAGATTAATGCGGTCCTGCTTCC GAACATTACATACTGTTTGTATATGTCGGCTGAAGACTTTGACGCTATTATAGCCGAGCACTTCCAG CCTGGGGATTGTGTTCTGGAACTGACATCGCGTCGTTTGATAAAAAGTGAGGACGACGCCATGGCT CTGACCCGCTTAATGATTCTGGAAGACTTAGGTGTGGACGCAGAGCTGTTGACGCTGATTGAGGCG GCTTTCGGCGAAATTTTCATCAATACATTTGCCCACTAAAATAAATTTAAATTCGGAGCCATGATGA AATCTGGAATGTTCTCACACTGTTTGTGAACACAGTCATTAACATTGTAATCGCAAGCAGAGTGTT GAGAGAACGGCTAACCGGATCACCATGTGCAGCATTATTGGAGATGACAATATCGTGAAAGGAG TCAAATCGGACAAATTAATGGCAGACAGGTGCGCCACCTGGTTGAATATGGAAGTCAAGATTATAG ATGCTGTGGTGGGCGAGAAAGCGCCTTATTTCTGTGGAGGGTTATTTTGTGTGACTCCGTGACCGG CACAGCGTGCCGTGTGGCAGACCCCTAAAAGGCTGTTTAAAGCTTGGCAAACCTCTGGCAGCAGA CGATGAACATGATGATGACAGGAGAAGGCATTCATGTAAGAGTCAACACGCTGGAACCGGATGG GTATTCTTTCAGAGCTGTGCAAGGCAGTAGAATCAAGGTATGAAACCGTAGGAACCTTCCATCATAG TTATGGCCATGACTACTCTAGCTAGCAGTGTTAAATCATTACAGCTACCTGAGAGGGGCCCTATAA CTCTCTACGGCTAACCTGAATGGACTACGACGTATCACGCCAAACATTTACAGCCGCGGTGTCAA AAACCGCGTGGACGTGGTTAACATCCCTGCTGGGAGGATCAGCCGTAATTATTATAAATTGGCTTGG TGCTGGCTACTATTGTGGCCATGTACGTGCTGACCAACCAGAAAACATAATTGAATACAGCAGCAAT TGGCAAGCTGCTTACATAGAATCGCGGCGATTGGCATGCCGCTTAAAAATTTTTATTTTTTTTC TTTTCTTTTCCGAATCGGATTTTGTTTTAAATTTTCAAAAAAAAAAAAAAAAAAAAAAAAAAAAA AA
VEE-UbAAY (SEQ ID NO:14); VEEV delivery vector with MHC class I mouse tumor epitopes SIINFEKL and AH1-A5 inserted
VEE-Luciferase (SEQ ID NO:15); VEEV 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 nulceic acid sequence (SEQ ID NO:46)
Tetanus toxoid amino acid sequence (SEQ ID NO:47)
PADRE nulceotide 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)>cGFP 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)
GPGPG linker (SEQ ID NO:56)
SAM <i>in vitro</i> transcription template DNA (SEQ ID NO:57); VEEV genome with nucleotides 7544-11175 deleted, plus minimal 5' T7-promoter (Bold Italic)
TAATACGACTCACTATAAT GggcggcgcatgagagaagccagaccaattacctacccaaaATGgagaagtgcacgttgacatcgaggaa gacagccattctcagagcttgcagcggagcttcccgcagttgaggtagaagccaagcaggtcactgataatgacctgctaagccagagcgttttcgcatctg gctcaaaactgatcgaacggaggtgacccatccgacacgatccttgacattggaagtgcgcccccgcagaatgtattcagcacaagtatcattgatctgt ccgatgagatgtcgggaagatccggacagattgtataagtatgcaactaagctgaagaaaaactgaaggaataactgataaggaattggacaagaaaatgaagg agctcggcggcctgatgagcaccctgacctgaaactgagactatgtcctccacgacgacgagtcgtgctgctacgaaggcgaagtcgctgtttaccaggatgt ataccggttgacggaccgacaagctctatcaccaagccaataagggagttagagtcgcctactggataggctttgacaccacctttatgtttaagaactggct ggagcatatccatcactctaccaactggcgccgacgaaccggttaacggctcgaacatagacccatgacagctctgacgtttggagcgggtcacgtagagggat gtccattctagaagaagtattgaaacctcaacaatgttctattctgttggctcgacctctaccacgagaagagggacttactgaggagctggcacctgccgt ctgtattcacttacgtggaagcaaaattacacatgtcgggtgagactatagttgctgcagcgggtacgtcgttaaaagaatagctatcagccaggcctgtatggg aagccttcaggctatgctgctacatgacccgcgagggattctgtgctgcaaaagtacagacacattgaacggggagagggctctttcccgtgtgacgtatgtg ccagctacattgtgaccaaactgactggcactggaacagatgctagtcggcagcgcgcaaaaactgctggtgggctcaaccagcgtatagtcgtaaacg gtcgcacccagagaacaccaataccatgaaaattacctttgcccgtagtgcccagcattgctaggtgggcaaaaggaataaaggaatcaagaagatga aaggccactaggactacgagatagacagttatgctatgggtgtgtggcctttagaagcacaagataactctattataagcggccggatacccaaacatcat caaagtgaacagcgtattccactatctgtgctgccagagatggcagtaaacattggagatcgggctgagaacaagaatcaggaaaattgtagaggacaca ggagccgtcacctctcattaccgccgaggacgtacaagaagctaaagtcgacggcagatgaggtcaagaggtgctgtaagccgaggaggtgctgagcagctca ccacctttggcagctgatgttgaggagcccactctggaagccgatgctgactgtgtacaagaggtggggccggctcagtgagacacctctgctgtgataaa ggttaccagctacgctggcaggacaagatcggctctacgctgtgctttccgcaggctgtactcaagagtgaataattctgtcatccacctctcgtgaaaca gcatagtataacacactctggccgaaaaggcgttatgccgtggaaccataccatggtaaagtagtgggtgccagagggacatgcaataaccctccaggactttc

agcacaagtatcatttatctgtccgatgagatgtcggagatccggacagattgtataagatgcaactaagctgaagaaaaactgtaaggaaataactgataagg
aattggacaagaaaatgaaggagctcggccgcatgagcgcacctgacctggaactgagactatgtccctccacgacgacgagtcgtcgtctacgaagg
caagtcgctgttaccaggtatgatacgcggtgacggaccgacaagctctatcaccagaacaaataggagtagatgcctactggtataggctttgacaccac
ccctttatgtttaaagaactgtgctggagcatalccatcactctaccaactggccgacgaaaccgtgttaacggctgtaacatagccctatgacgctctgacgttat
ggagcggcactgtagggatgtccattctagaagaagtattgaaaccatccaacaatgttctattctgttggctcaccatctaccacgagaagaggactta
ctgaggagctggcactcctgtctgtattcacttacgtgcaagcaaaatfacacatgtcgggtgtagactatagttgtagcagcgggtacgtcgttaaaagaatag
ctatcagtcaggcctgtatgggaagcctcaggctatgctgctacgatgaccgcgagggatcttctgtcgtcgaagtgacagacacattgaacggggagagggt
ctctttcccggtgacgctatgtccagctacattgtgtgacaaatgactggcactggaacagatgtagtgcggacgacgcgcaaaaactgctggtgggct
caaccagctatagtcgtaacggctgacccagagaaacaccaatacatgaaaaatfacccttggccgtatgtggccaggcattgtaggtgggcaaaaggaat
ataaggagatcaagaagatgaaaggccactagactacgagatagacagttagctatggggtgtgtggcctttagaaggcacaagataactctattataagc
gcccggatcccaaacatcatcaaatgaaacgcgattccactcattctgctgccaggataggcagtaaacattggagatcgggctgagaacaagaatcag
gaaaatgtagaggagcacaaggagccgtcacctctaccgccgagagctacaagaagctaatgtcgcagccgatgaggtaaggaggtgctggaagcc
gaggagttgcgcgacgtctaccaccttggcagctgatgttaggagccactctggaagccgatgctgactgatgttacaagagcctgggcccggctcagtg
agacacctgtgctgataaaggctaccagctacgctggcaggacaagatcggctctacgctgtgcttctccgaggtctactcaagatgaaaaatlatctg
catccaccctctcgtgaacaagctatgataaacacactctgcccgaaaaaggcgttatgccgtggaaccataccatgtaagtagtggcagaggacat
gcaatccccctcaggacttcaagctctgagtgaagtgccaccattgtgtaaacgaacgtgagttcgttaaacaggtacctgacacatattgccaacatgagg
agcgtgaaactgatgaagaatattcaaaaactgcaagccccgagcagcagcggcaatacctgtacgacatcgcaggaacagtgctcaagaagaac
tagtactgggctaggcctcacaggcagctggtgatctccctccatgaattcgcctacgagagctgagaacacgaccagccgctcttaccagtaaccaac
cataggggtgtatggcgtccaggatcaggcaagctggcacttaaaagcgcagctacaaaaaagatctagtggtagcgaagaagaaaaactgtgcaga
aattataaggacgtcaagaaaatgaaaggctggacgtcaatgccagaactgtggactcagctcttgaatggatgcaaacaccccgtagagacctgtatattg
acgaagctttgctgcatgacgtactctcagagcgtcagaccatataagacctaanaaggcagctgctctcggggtacccaaacagtcgggttttitaacatg
atgtccgtaaaatgcattttaaccacgagattgacacaaagcttccacaaaagcactctcggcgtgactaaactgtgactcggctgctcaacctgttttacg
acaaaaaatgagaacgacgaatccgaaagagactaagattgattgacactaccggcagtagcaaacctaaagcagacatctactcactgtttcagagg
tgggtgaagcagttgcaaatgattacaaggaacgaataatgacggcagctgctctcaaggcgtgaccgttaaagggtgtatgccgttcggtacaaggtga
atgaaatcctctgacgacccacctcagaacatgtaacgtcctactgaccgcagggagaccgatcgtgtgaaaacactagccggcgaccatggataa
aaactgactgccaagtacctgggaattcactgccacgatagaggagtggcaagcagagcatgatccatcatgaggacacatctggagagaccggacccta
ccgacgtctccagaataaggcaaacgtgtgtggccaaggctttagtccggctgctgaagaccgctggcagacatgaccactgaacaatgaaactgtgga
ttatftgaaacggacaagctcactcagcagagatagattgaaccaactatgctgtgaggttcttggactcgtctggactccggtctatttctgaccacctgtccg
ttatccattaggaataactcactgggataactccccgtcgcctaactgtacggcgtcaataaagaagtggtccgtcagctctctcgcaggtaaccacaactgcctcgg
gcagttgccactggaagagctatgacatgaacactggtacactgcgcaattatgaccgcgcaataacctgactctgaaacagaagactgcctcatgctttatgctc
tccaccataatgaaccccacagagtgactttctcaltcgtcagcaattgaaggcgaactgtcctggtgctggggaaggtgtccgtcccaggcaaaatgg
ttgactggtgtgacagccgctgaggtcactctcagagctcggctggatttaggcatcccagggtgatgtcccaaatatgacataatattgtaatgtgaggacccc
ataataatcactcactatcagcagtgtaagaccatgccattagcttagatgttgaccaagaagctgtgctgactgaatcccggcgaacctgtgtcagata
ggtattggttactgacagggccagcgaagcaltggtgctatagcggcagttcaagtttccgggtatgcaaacgaaalccacttgaagagacgga
agttctgtttgattcattgggtacgatcgaaggcccgtacgcacaactctacaagctttcatcaacttgaccaacattatatacaggttccagactccacgaagccg
gatgtgaccctcatatcatgtgtgaggggatattgccacggccaccgaaggagtattataaatgctgtaacagcaaaagcaacctggcggaggggtgt
ggcggagcgtgtataaagaattccggaaagcttcgattacagccgatcgaagtaggaaaagcgcgactggtcaaaaggtgcagtaaacatatacatcatgccgta
ggaccaaacctcaacaagttcggaggtgaaaggtgacaacagttggcagagctttagtccatcgtlaagattgcaacgatacaaatcaagtagcagta
gattccactgttccaccggcactcttccgggaacaagatcgaactaaccaatcattgaaccatttgcagacgtttagacaccactgacagatgtagccat
actgcagggacaagaatgggaatgactctcaaggaaagcagtggttaggagagaagcagtgaggagatgacatccgacgactctcagtgacagaacct
gatgcagagctggtgaggtgcatccgaagattcttggctggaaggaaaggctacagcacaagcagtgcaaaaacttctcatatttgaagggaccaagttca
ccaggcggccaaggatagcagaatataatgccatgtggccgttgaacggaggccaatgagcaggtatgcatatctcggagaagacatgagcagttat
aggtcgaatgccccgtcgaagagcggaaagcctccacaccactagcacgctgcttctgtgcatccatgccatgactccagaagagtagcgcctaaaa
gcctcagctcagaacaatfactgtgtgctcatcttccattgccgaagtataaactggtgtgcaagaatccatgctcccagcctatattgtctaccgaaa
gtcctgctgataattatcaaggaaatgctcgtgaaacaccaccggtagacgagactcggagccatcggcagagaaccaatccacagaggggacacctga
acaaccaccactataaccgaggatgagaccagactagaacgctgagccgatcatcatcgaagaggaagaagagatgataaattgtctgcatgagccc
gaccaccaggtgctgcaagtcgagcagacattcacggccgcccctgtatctagctcatctggtccattctcatgcatccgactttgatgtggacgattatcc
atactgacacctggaggagctagcgtgaccagcgggcaacgtcagccgagactaactcttactcgcaaaagatgaggttctgctgcgcgaccggctgct
cgctcgaacagatcaggaaacctccacatcccgtccgcgcaagaacaccgtcactgacccagcagggcctgctcagaaccagcctggtctcaaccgg
ccaggcgttaaatagggtgattacaagagaggattgaggcgttgtagcaacaacaatgacgggtttagcgggtgcatacatctttccgacaccggca
agggcatttacaacaataatgataaggcaaacgtgctatccgaagtgtgtggagaggaccgaattggagatttctgatccccgcctcaccagaaaaa
gaagaattactacgcaagaataactgtaaatccacacctgtaacagaagcagataaccagtcaggaaggtggagaacatgaagccataacagctagacgt
attctgcaaggcctaggcattattgaaagcagaaggaaggtggagtgtaccgaacctgcatctgtctcttgtattcatctagtgtgaaccgtgcttttcaag
cccaaggtcagtggaagcctgtaacgccatgtgaaagagaacttccgactgtggtcttactgtatttccagagtacgatgctatttggacatggttgacg
gagcttcatgctcttagactgcccgttttccctgcaagctcgcgacttccaagaacactctatttgaaccacaatacagtcggcagtgctctcagc
gatccagaacagcctcagaacgtcctggcagctccacaaaaagaatgcaatgtcacgcaaatgagagaattgccctattgattcggcggcctttaaagtg
aatgttcaagaatgctgtgataatgaaatgtggaagctttaaagaaaacccatcagcgttactgaagaaaactggtgaaatfacattaccaatataaagg
acaaaaactgctgctcttttgcgaagacataatgaaatgtgtaggacatacaatgacaggttgaatggactttaaagagagactgaaagtactcca
ggaacaaaactactgaagaacggccaaaggtacaggtgatccaggctgccgatccgtagcaacagcgtatctgtcggaaatccaccgagagctggttaggag
attaatgctgctcttccgaacattcacaactgttgatgtcggctgaagactttgacgctattatagccgagcactccagcctggggattgtgtctgaaact
gacatcgcgtgtttgataaaagtgaggacgacgccatggctctgaccgcgttaatgattctggaagacttaggtgtagcgcagagctgtgacgctgattgaggc

```
ggcttfcggcgaattcatcaatacattgccactaaaactaaatftaaattcggagccatgatgaaatctggaatgttcctcacactgtttgtaacacagtcattaac  
attgtaatcgaagcagaggtgtgagagaacggctaaccggatcacatgtgcagcattcattggagatgacaatcgtgaaaggagcacaacggacaaataat  
ggcagacaggtgcgccacctgggtgaatatggaagcaagattatagatgctgtggtggcgagaaagcgcctatttctgagggtttatttgtgtgactccgtg  
accggcacagcgtgccgtgtggcagacccccaaaaaggctgttaagcttggcaaacctciggcagcagacgatgaacatgatgatgacaggagaaggcattg  
catgaagagcaacacgctggaaccgagtggtattcttcagagctgtgcaaggcagtagaatcaaggtatgaaaccgtagggaacttccatcatagttatggccatg  
actactctagctagcaggttaaatcattcagctacctgagaggggccctataactctctacggcTA AcctgaaaggactacgactTatcacgccaacatttac  
agccgcggtgtcaaaaaccgctggacgtggttaacatccctgctgggagatcagccgtaattattataattggcttggctgctgctactattgtggccatgtacgtg  
ctgaccaaccagaacataattgaatcacgagcaattgcaagctgctfacatagaactcgcggcgattggcatgccgcctaaaatTTTTTTTTTTTTTTTTTTT  
cgaatcgattttgttttaatttcAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA  
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA  
AAAAAAAAAAAtacgtagttaaac
```

Equivalents and Scope

[00286] In the claims, articles such as “a,” “an,” and “the” may mean one or more than one unless indicated to the contrary or otherwise evident from the context. Claims or descriptions that include “or” between one or more members of a group are considered satisfied if one, more than one, or all of the group members are present in, employed in, or otherwise relevant to a given product or process unless indicated to the contrary or otherwise evident from the context. The present disclosure includes embodiments in which exactly one member of the group is present in, employed in, or otherwise relevant to a given product or process. The present disclosure includes embodiments in which more than one, or all of the group members are present in, employed in, or otherwise relevant to a given product or process.

[00287] Furthermore, the present disclosure encompasses all variations, combinations, and permutations in which one or more limitations, elements, clauses, and descriptive terms from one or more of the listed claims is introduced into another claim. For example, any claim that is dependent on another claim can be modified to include one or more limitations found in any other claim that is dependent on the same base claim. Where elements are presented as lists, e.g., in Markush group format, each subgroup of the elements is also disclosed, and any element(s) can be removed from the group. It should be understood that, in general, where the present disclosure, or aspects of the present disclosure, is/are referred to as comprising particular elements and/or features, certain embodiments of the present disclosure or aspects of the present disclosure consist, or consist essentially of, such elements and/or features. For purposes of simplicity, those embodiments have not been specifically set forth in haec verba herein. It is also noted that the terms “comprising” and “containing” are intended to be open and permits the inclusion of additional elements or steps. Where ranges are given, endpoints are included. Furthermore, unless otherwise indicated or otherwise evident from the context and understanding of one of ordinary skill in the art, values that are expressed as ranges can assume any specific value or sub-range within the stated ranges in different embodiments of the present disclosure, to the tenth of the unit of the lower limit of the range, unless the context clearly dictates otherwise.

[00288] This application refers to various issued patents, published patent applications, journal articles, and other publications, all of which are incorporated herein by reference. If there is a conflict between any of the incorporated references and the instant specification, the specification shall control. In addition, any particular embodiment of the present disclosure that falls within the prior art may be explicitly excluded from any one or more of the claims. Because such embodiments are deemed to be known to one of ordinary skill in the art, they may be excluded even if the exclusion is not set forth explicitly herein. Any particular embodiment of

the present disclosure can be excluded from any claim, for any reason, whether or not related to the existence of prior art.

[00289] Those skilled in the art will recognize or be able to ascertain using no more than routine experimentation many equivalents to the specific embodiments described herein. The scope of the present embodiments described herein is not intended to be limited to the above Description, but rather is as set forth in the appended claims. Those of ordinary skill in the art will appreciate that various changes and modifications to this description may be made without departing from the spirit or scope of the present disclosure, as defined in the following claims.

CLAIMS

What is claimed is:

1. A composition for delivery of a self-amplifying expression system comprising a single-stranded RNA (ssRNA) vector, wherein the ssRNA vector comprises a 7-methylguanylate (m⁷G cap), a polyadenylation (polyA) tail, a self-amplifying backbone, and one or more modified nucleosides, wherein the one or more modified nucleosides comprise m⁵C and no other modified nucleotides, optionally other than an m⁷G cap analog.
2. A composition for delivery of a self-amplifying expression system comprising a single-stranded RNA (ssRNA) vector, wherein the ssRNA vector comprises a 7-methylguanylate (m⁷G cap), a polyadenylation (polyA) tail, a self-amplifying backbone, and one or more modified nucleosides.
3. The composition of claim 1 or 2, wherein the self-amplifying backbone comprises a polynucleotide selected from a self-replicating RNA virus.
4. The composition of any one of claims 1-3, wherein the self-replicating RNA virus is selected from the group comprising an alphavirus, a flavivirus, a measles virus, and a rhabdovirus.
5. The composition of any one of claims 2-4, wherein the one or more modified nucleosides comprise methyl-5-cytosine (m⁵C), methyl-6-adenosine (m⁶A), ribose-methylation (2'-O-Me), s-thiouridine, (s²U), 5-methyluridine (m⁵U), N1-methyl pseudouridine (m¹Ψ), pseudouridine (Ψ), or any combination thereof.
6. The composition of any one of claims 2-4, wherein the one or more modified nucleosides comprise m⁵C.
7. The composition of any one of claims 2-4, wherein the one or more modified nucleosides comprise m⁵C and no other modified nucleotides, optionally other than an m⁷G cap analog.
8. The composition of any of the above claims, wherein the ssRNA vector is purified.
9. The composition of claim 8, wherein the ssRNA vector is purified by chromatography.
10. The composition of claim 9, wherein the chromatography comprises a cellulose chromatography system or an affinity-based separation system.
11. The composition of claim 10, wherein the affinity-based separation system is a deoxythymidine (dT) oligonucleotide (Oligo(dT)) system.

12. The composition of any of the above claims, wherein greater than 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, or 100% of adenine, guanine, cytidine, and/or uridine nucleosides of the ssRNA vector are modified nucleosides.
13. The composition of any of the above claims, wherein greater than 25% of adenosine, guanosine, cytidine, and/or uridine nucleosides of the ssRNA vector comprise modified nucleosides.
14. The composition of claim 13, wherein 100% of adenine, guanine, cytidine, and/or uridine nucleosides of the ssRNA vector comprise modified nucleosides.
15. The composition of any of the above claims, wherein the ssRNA vector comprises 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, or 100% of cytidine nucleosides comprise modified nucleosides.
16. The composition of any of the above claims, wherein greater than 25% of cytidine nucleosides of the ssRNA vector comprise modified nucleosides.
17. The composition of any of the above claims, wherein the ssRNA comprises 100% of cytidine nucleosides are modified nucleosides.
18. The composition of any of the above claims, wherein the composition comprises less than 30%, less than 25%, less than 20%, less than 15%, less than 10%, less than 5%, less than 2%, less than 1%, less than 0.5%, or less than 0.1% of a contaminant.
19. The composition of claim 18, wherein the contaminant comprises salt, detergent, and/or double stranded RNA (dsRNA).
20. The composition of claim 18, wherein the contaminant comprises dsRNA.
21. The composition of any of the above claims, wherein the ssRNA vector comprises 70% or more, 80% or more, 85% or more, 90% or more, or 95% or more of the total RNA present in the composition.
22. The composition of any of the above claims, wherein the ssRNA vector comprises 97% or more, 98% or more, 99% or more, or 99.5% or more of the total RNA present in the composition.
23. The composition of any of the above claims, wherein the ssRNA vector is the only RNA species present in the composition.

24. The composition of any one of claims 18-23, wherein the total RNA, dsRNA, and/or ssRNA content is assessed by dot blot or ELISA, optionally wherein the dot blot or ELISA assessment comprises detection with an antibody specific to RNA, dsRNA, and/or ssRNA.
25. The composition of any one of claims 18-23, wherein the total RNA, dsRNA, and/or ssRNA content is assessed by capillary electrophoresis.
26. The composition of any one of claims 18-23, wherein the total RNA, dsRNA, and/or ssRNA content is assessed liquid chromatography.
27. The composition of any one of claims 24 or 26, wherein the RNA, dsRNA, and/or ssRNA content is quantified as area under the curve (AUC).
28. The composition of any of the above claims, wherein the ssRNA vector is produced by *in vitro* transcription.
29. The composition of any of the above claims, wherein the m⁷G cap comprises an m⁷G cap analog.
30. The composition of claim 29, wherein the m⁷G cap analog comprises a trinucleotide m⁷G-ppp-A-U cap analog or dinucleotide m⁷G-ppp-A cap analog.
31. The composition of any of the above claims, wherein the ssRNA vector comprises a cassette comprising at least one nucleic acid sequence for delivery, optionally wherein the at least one nucleic acid sequence comprises a polypeptide-encoding nucleic acid sequence, optionally wherein the polypeptide-encoding nucleic acid sequence is an antigen-encoding nucleic acid sequence, and wherein the cassette is operably linked to or operably inserted into the self-amplifying backbone.
32. The composition of any of the above claims, wherein the self-amplifying backbone comprises a polynucleotide selected from a self-replicating RNA virus comprising at least one nucleic acid sequence, from 5' to 3', described by the formula:
m⁷G-ppp-N₁-N₂-N_V, wherein
m⁷G is a 7-methylguanylate (m⁷G) cap,
ppp is a triphosphate bridge,
N₁ is a first nucleotide of the self-amplifying backbone corresponding to a first endogenous 5' nucleotide of the self-replicating RNA virus,
N₂ is a second nucleotide of the self-amplifying backbone corresponding to a second endogenous 5' nucleotide of the self-replicating RNA virus, and

Nv comprises (1) one or more additional nucleic acid sequences of the self-amplifying backbone, and (2) a cassette comprising at least one nucleic acid sequence for delivery, optionally wherein the at least one nucleic acid sequence comprises a polypeptide-encoding nucleic acid sequence, optionally wherein the polypeptide-encoding nucleic acid sequence is an antigen-encoding nucleic acid sequence, and wherein the cassette is operably linked to or operably inserted into the self-amplifying backbone.

33. The composition of any of the above claims, wherein the composition further comprising a nanoparticle delivery vehicle.

34. The composition of claim 33, wherein the nanoparticle delivery vehicle is a lipid nanoparticle (LNP).

35. A pharmaceutical composition comprising the composition of any of the above claims and a pharmaceutically acceptable carrier.

36. A method for treating a subject with a disease, the method comprising administering to the subject an amount of the composition of any of the above composition claims or the pharmaceutical composition of claim 35.

37. The method of claim 36, wherein the disease is cancer or an infectious disease.

38. The method of claim 37, wherein the infectious disease is caused by a virus selected from the group comprising, HPV, influenza, TB, CMV, HMPV, PIV, chikungunya virus, Zika virus, SARS-CoV-2, and pan-coronavirus.

39. The method of any of claims 36-38, wherein the amount of the ssRNA vector in the composition is 1000 µg or less, 100 µg or less, 50 µg or less, 30 µg or less, 10 µg or less, 5 µg or less, or 1 µg or less.

40. The method of any of claims 36-39, wherein there is a reduction an innate immune response in the subject compared to a control subject receiving a composition comprising a nucleic acid sequence comprising no modified nucleosides.

41. The method of claim 40, wherein the innate immune response is an IFN response.

42. The method of claim 41, wherein the IFN response is IRF-3 expression, IRF-7 expression, or a combination thereof.

43. The method of any of claims 36-42, wherein there is an increase in replication of the ssRNA vector compared to a control composition comprising an otherwise identical ssRNA vector but comprising no modified nucleosides.

44. A method of purifying a self-amplifying expression system from a nucleic acid mixture, the method comprising purifying the self-amplifying expression system through a cellulose chromatography system or an affinity-based separation system, wherein the self-amplifying expression system comprises a single-stranded RNA (ssRNA) vector, and wherein the ssRNA vector comprises an m⁷G cap, a polyA tail, and a self-amplifying backbone.
45. A method of reducing double stranded RNA (dsRNA) in a nucleic acid mixture, the method comprising:
- (A) purifying the self-amplifying expression system through a cellulose chromatography system or an affinity-based separation system; and/or
 - (B) producing the self-amplifying expression system such that the self-amplifying expression system comprises modified nucleosides; and
- wherein the self-amplifying expression system comprises a single-stranded RNA (ssRNA) vector, wherein the ssRNA vector comprises an m⁷G cap, a polyA tail, and a self-amplifying backbone.
46. The method of claim 44 or 45, wherein the affinity-based separation system comprises a deoxythymidine (dT) oligonucleotide (Oligo(dT)) system.
47. The method of claim 44, wherein the ssRNA vector comprises a modified nucleoside.
48. The method of any one of claims 45-47, wherein the modified nucleoside comprises methyl-5-cytosine (m5C), methyl-6-adenosine (m6A), ribose-methylation (2'-O-Me), s-thiouridine, (s2U), 5-methyluridine (m5U), N1-methyl pseudouridine (m1Ψ), pseudouridine (Ψ), or any combination thereof.
49. The method of any one of claims 45-47, wherein the modified nucleoside consists of methyl-5-cytosine (m5C).
50. A composition mixture, wherein the composition comprises a self-amplifying expression system comprising:
- (a) a single-stranded RNA (ssRNA) vector, wherein the ssRNA vector comprises an m⁷G cap, a polyadenylation (polyA) tail, and a self-amplifying backbone; and
 - (b) double stranded RNA (dsRNA), and
- wherein the dsRNA comprises less than 20%, less than 10%, less than 5%, less than 2%, less than 1%, less than 0.5%, or less than 0.1% of the total RNA present in the mixture.

51. A composition mixture, wherein the composition comprises a self-amplifying expression system comprising:
- (a) a single-stranded RNA (ssRNA) vector, wherein the ssRNA vector comprises an m⁷G cap, a polyadenylation (polyA) tail, and a self-amplifying backbone; and
 - (b) double stranded RNA (dsRNA), and
- wherein the ssRNA vector comprises 70% or more, 80% or more, 85% or more, 90% or more, 95% or more, 97% or more, 98% or more, 99% or more, or 99.5% or more of the total RNA present in the mixture.
52. The mixture of claim 50 or 51, wherein the only detectable RNA comprises the ssRNA vector.
53. The mixture of any of the above claims, wherein the total RNA, dsRNA, and/or ssRNA content is assessed by dot blot or ELISA, optionally wherein the dot blot or ELISA assessment comprises detection with an antibody specific to RNA, dsRNA, and/or ssRNA.
54. The mixture of any of the above claims, wherein the total RNA, dsRNA, and/or ssRNA content is assessed by capillary electrophoresis.
55. The mixture of any of the above claims, wherein the total RNA, dsRNA, and/or ssRNA content is assessed liquid chromatography.
56. The composition of any one of claims 53-55, wherein the RNA, dsRNA, and/or ssRNA content is quantified as area under the curve (AUC).
57. The mixture of any of the above claims, wherein the ssRNA is produced by *in vitro* transcription.
58. The mixture of any of the above claims, wherein the m⁷G cap comprises an m⁷G cap analog.
59. The mixture of claim 58, wherein the m⁷G cap analog comprises a trinucleotide m⁷G-ppp-A-U cap analog or dinucleotide m⁷G-ppp-A cap analog.
60. The mixture of any of the above claims, wherein the ssRNA vector comprises a cassette comprising at least one nucleic acid sequence for delivery, optionally wherein the at least one nucleic acid sequence comprises a polypeptide-encoding nucleic acid sequence, optionally wherein the polypeptide-encoding nucleic acid sequence is an antigen-encoding nucleic acid sequence, and wherein the cassette is operably linked to or operably inserted into the self-amplifying backbone.

61. The mixture of any of the above claims, wherein the self-amplifying backbone comprises a polynucleotide selected from a self-replicating RNA virus comprising at least one nucleic acid sequence, from 5' to 3', described by the formula:

m^7G -ppp- N_1 - N_2 - N_v , wherein

m^7G is a 7-methylguanylate (m^7G) cap,

ppp is a triphosphate bridge,

N_1 is a first nucleotide of the self-amplifying backbone corresponding to a first endogenous 5' nucleotide of the self-replicating RNA virus,

N_2 is a second nucleotide of the self-amplifying backbone corresponding to a second endogenous 5' nucleotide of the self-replicating RNA virus, and

N_v comprises (1) one or more additional nucleic acid sequences of the self-amplifying backbone, and (2) a cassette comprising at least one nucleic acid sequence for delivery, optionally wherein the at least one nucleic acid sequence comprises a polypeptide-encoding nucleic acid sequence, optionally wherein the polypeptide-encoding nucleic acid sequence is an antigen-encoding nucleic acid sequence, and wherein the cassette is operably linked to or operably inserted into the self-amplifying backbone.

62. The mixture of any of the above claims, wherein the ssRNA vector comprises one or more modified nucleosides.

63. The mixture of claim 62, wherein the one or more modified nucleosides comprise methyl-5-cytosine (m5C), methyl-6-adenosine (m6A), ribose-methylation (2'-O-Me), s-thiouridine, (s2U), 5-methyluridine (m5U), N1-methyl pseudouridine (m1Ψ), pseudouridine (Ψ), or any combination thereof.

64. The mixture of claim 62, wherein the one or more modified nucleosides comprise m5C.

65. The mixture of claim 62, wherein the one or more modified nucleosides comprise m5C and no other modified nucleotides, optionally other than a cap analog.

66. The mixture of any of the above claims, wherein the ssRNA vector is purified.

67. The mixture of claim 66, wherein the ssRNA vector is purified by chromatography.

68. The mixture of claim 67, wherein the chromatography comprises a cellulose chromatography system or an affinity-based separation system.

69. The mixture of claim 68, wherein the affinity-based separation system is a deoxythymidine (dT) oligonucleotide (Oligo(dT)) system

70. The composition, mixture, or method of any one of the above claims,

wherein the composition for delivery of the self-amplifying expression system comprises:

(A) the self-amplifying expression system, wherein the self-amplifying expression system comprises one or more self-amplifying mRNA (SAM) vectors, wherein the one or more SAM vectors comprise:

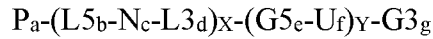
(a) the self-amplifying backbone, wherein the self-amplifying backbone comprises the nucleic acid sequence set forth in SEQ ID NO:6, wherein the self-amplifying backbone sequence comprises a subgenomic promoter nucleotide sequence and a poly(A) sequence, wherein the subgenomic promoter sequence is endogenous to the self-replicating RNA virus, wherein the poly(A) sequence is endogenous to the self-replicating RNA virus backbone; and

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

- a. an epitope-encoding nucleic acid sequence, optionally comprising: (1) at least one alteration that makes the encoded epitope sequence distinct from the corresponding peptide sequence encoded by a wild-type nucleic acid sequence, or (2) 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) optionally, a lipid-nanoparticle (LNP), wherein the LNP encapsulates the self-amplifying expression system.

71. The composition, mixture, or method of any one of the above claims, wherein an ordered sequence of each element of the cassette in the composition for delivery of the self-amplifying expression system is described in the formula, from 5' to 3', comprising



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.

72. The composition of claim 71, wherein for each X the corresponding N_c is a distinct MHC class I epitope-encoding nucleic acid sequence.

73. The composition of claim 71 or 72, wherein for each Y the corresponding U_f is a distinct MHC class II epitope-encoding nucleic acid sequence.

74. The composition of any one of claims 71-73, wherein

$$a = 0, b = 1, d = 1, e = 1, g = 1, h = 1, X = 10, Y = 2,$$

the at least one promoter nucleotide sequence is a single subgenomic promoter nucleotide sequence provided by the self-amplifying backbone,

the at least one polyadenylation poly(A) sequence is a poly(A) sequence of at least 80 consecutive A nucleotides provided by the self-amplifying backbone,

the cassette is integrated between the subgenomic promoter nucleotide sequence and the poly(A) sequence, wherein the cassette is operably linked to the subgenomic 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 self-amplifying 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.

75. The composition, mixture, or method of any one of the above claims, wherein the at least one nucleic acid sequence for delivery comprises the polypeptide-encoding nucleic acid sequence.
76. The composition of claim 75, wherein the polypeptide-encoding nucleic acid sequence encodes the antigen-encoding nucleic acid sequence.
77. The composition of claim 76, wherein the antigen-encoding nucleic acid sequence comprises a MHC class I epitope, a MHC class II epitope, an epitope capable of stimulating a B cell response, or a combination thereof.
78. The composition of claim 76 or 77, wherein the antigen-encoding nucleic acid sequence comprises sequence encoding a full-length protein, a protein subunit, a protein domain, or a combination thereof.
79. The composition of claim 75, wherein the polypeptide-encoding nucleic acid sequence encodes a full-length protein or functional portion thereof.
80. The composition of claim 79, wherein the full-length protein or functional portion thereof is selected from the group consisting of: an antibody, a cytokine, a chimeric antigen receptor (CAR), a T-cell receptor, and a genome-editing system nuclease.

81. The composition, mixture, or method of any one of the above claims, wherein the at least one nucleic acid sequence for delivery comprises at least one nucleic acid sequence comprising a non-coding nucleic acid sequence.
82. The composition of claim 81, wherein the non-coding nucleic acid sequence is an RNA interference (RNAi) polynucleotide or genome-editing system polynucleotide.
83. 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.
84. 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.
85. The composition of claim 83 or 84, wherein the ionizable amino lipids comprise MC3-like (dilinoleylmethyl-4-dimethylaminobutyrate) molecules.
86. The composition of any of the above claims, wherein the LNP-encapsulated expression system has a diameter between 60-140nm.
87. The composition of any of the above claims, wherein the composition for delivery of the self-amplifying expression system is formulated for intramuscular (IM), intradermal (ID), subcutaneous (SC), intravitreal (IVT), intrathecal, or intravenous (IV) administration.
88. The composition of any of the above claims, wherein the composition for delivery of the self-amplifying expression system is formulated for intramuscular (IM) administration.
89. The composition, mixture, or method of any one of the above claims, wherein the cassette is integrated between the at least one promoter nucleotide sequence and the at least one poly(A) sequence.
90. The composition, mixture, or method of any one of the above claims, wherein the at least one promoter nucleotide sequence is operably linked to the cassette.
91. The composition, mixture, or method of any one of the above claims, wherein the ssRNA vector comprises a positive-stranded RNA vector.
92. The composition, mixture, or method of any one of the above claims, wherein the ssRNA vector comprises a negative-stranded RNA vector.
93. The composition of claim 92, wherein the negative-stranded RNA vector comprises at least one polynucleotide sequence of a measles virus or a rhabdovirus.

94. The composition, mixture, or method of any one of the above claims, wherein the ssRNA vector is self-amplifying within a mammalian cell.
95. The composition, mixture, or method of any one of the above claims, wherein the self-replicating RNA virus is selected from the group consisting of: an alphavirus; a flavivirus, a measles, and a rhabdovirus.
96. The composition, mixture, or method of any one of the above claims, wherein the self-amplifying backbone comprises at least one polynucleotide sequence of an alphavirus, optionally wherein the alphavirus is selected from the group consisting of: Aura virus, a Fort Morgan virus, a Venezuelan equine encephalitis virus, a Ross River virus, a Semliki Forest virus, a Sindbis virus, and a Mayaro virus.
97. The composition, mixture, or method of any one of the above claims, wherein the self-amplifying backbone comprises at least one nucleotide sequence of a Venezuelan equine encephalitis virus.
98. The composition of claim 96 or 97, wherein the self-amplifying backbone comprises at least sequences for nonstructural protein-mediated amplification, a subgenomic 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.
99. The composition of claim 96 or 97, wherein the self-amplifying backbone comprises at least sequences for nonstructural protein-mediated amplification, a subgenomic 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.
100. The composition of claim 98 or 99, 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.
101. The composition of any one of claims 98-100, wherein the self-amplifying backbone does not encode structural virion proteins capsid, E2 and E1, optionally wherein E1 is a full-length E1, or does not encode structural virion proteins Capsid, E3, E2, 6K.

102. The composition of claim 101, wherein the cassette is inserted in place of structural virion proteins within the polynucleotide 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.
103. The composition of claim 96 or 97, wherein the Venezuelan equine encephalitis virus comprises the sequence of SEQ ID NO:3 or SEQ ID NO:5.
104. The composition of claim 96 or 97, 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.
105. The composition of claim 104, wherein the self-amplifying backbone comprises the sequence set forth in SEQ ID NO:6 or SEQ ID NO:7.
106. The composition of claim 104 or 105, wherein the cassette is inserted at position 7544 in SEQ ID NO:6 or SEQ ID NO:7.
107. The composition of claim 102-106, 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 for delivery are in separate open reading frames.
108. The composition, mixture, or method of any one of the above claims, wherein the ssRNA vector comprises at least one promoter nucleotide sequence.
109. The composition, mixture, or method of any one of claim 108, wherein the at least one promoter nucleotide sequence comprises the native promoter nucleotide sequence encoded by the self-replicating RNA virus, optionally wherein the native promoter nucleotide sequence is a subgenomic promoter nucleotide sequence.
110. The composition, mixture, or method of any one of claim 108, wherein the at least one promoter nucleotide sequence is an exogenous RNA promoter.
111. The composition, mixture, or method of any one of the above claims, wherein the ssRNA vector comprises a second promoter sequence, optionally wherein the second promoter nucleotide sequence is a subgenomic promoter nucleotide sequence.
112. The composition, mixture, or method of claim 111, wherein the second promoter nucleotide sequence comprises multiple subgenomic promoter nucleotide sequences, wherein each subgenomic promoter nucleotide sequence provides for transcription of one or more of the separate open reading frames.

113. The composition, mixture, or method of any one of the above claims, wherein the at least one antigen-encoding nucleic acid sequence comprises two or more antigen-encoding nucleic acid sequences.
114. The composition of claim 113, wherein each antigen-encoding nucleic acid sequence is linked directly to one another.
115. The composition, mixture, or method of any one of the above claims, 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.
116. The composition of claim 115, 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.
117. The composition of claim 116, 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.
118. The composition of claim 116, wherein the linker comprises 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.
119. The composition of claim 115, 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.
120. The composition of claim 119, wherein the linker comprises the sequence GPGPG.
121. The composition, mixture, or method of any one of the above claims, 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 epitope-encoding nucleic acid sequence.
122. The composition of claim 121, 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.

123. The composition, mixture, or method of any one of the above claims, 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.

124. The composition, mixture, or method of any one of the above claims, 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.

125. The composition, mixture, or method of any one of the above claims, 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.

126. The composition, mixture, or method of any one of the above claims, 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.

127. The composition, mixture, or method of any one of the above claims, wherein each antigen-encoding nucleic acid sequence independently comprises at least 2-10, 2, 3, 4, 5, 6, 7, 8, 9, or 10 epitope-encoding nucleic acid sequences, optionally wherein each epitope-encoding nucleic acid sequence encodes a distinct epitope-encoding nucleic acid sequence.

128. The composition, mixture, or method of any one of the above claims, wherein each antigen-encoding nucleic acid sequence independently 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 epitope-encoding nucleic acid sequences, optionally wherein each epitope-encoding nucleic acid sequence encodes a distinct epitope-encoding nucleic acid sequence.

129. The composition, mixture, or method of any one of the above claims, wherein each antigen-encoding nucleic acid sequence independently 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 epitope-encoding nucleic acid sequences.

130. The composition, mixture, or method of any one of the above claims, wherein each antigen-encoding nucleic acid sequence independently comprises at least 2-400 epitope-encoding nucleic acid sequences and wherein at least two of the epitope-encoding nucleic acid sequences encode epitope sequences or portions thereof that are presented by MHC class I on a cell surface.

131. The composition of 74, wherein at least two of the MHC class I epitopes are presented by MHC class I on a cell surface, optionally a tumor cell surface or an infected cell surface.

132. The composition, mixture, or method of any one of the above claims, 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.

133. The composition of any one of claims 71-73, or 83-132, wherein the at least one MHC class II epitope-encoding nucleic acid sequence is present.

134. The composition of any one of claims 71-73, or 83-132, 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.

135. The composition, mixture, or method of any one of the above claims, 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.

136. The composition, mixture, or method of any one of the above claims, 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.

137. The composition, mixture, or method of any one of the above claims, wherein the at least one promoter nucleotide sequence or the second promoter nucleotide sequence is inducible.

138. The composition, mixture, or method of any one of the above claims, wherein the at least one promoter nucleotide sequence or the second promoter nucleotide sequence is non-inducible.

139. The composition, mixture, or method of any one of the above claims, wherein the poly(A) sequence comprises a poly(A) sequence native to the self-replicating virus.

140. The composition, mixture, or method of any one of the above claims, wherein the at least one poly(A) sequence comprises a poly(A) sequence exogenous to the self-replicating virus.

141. The composition, mixture, or method of any one of the above claims, wherein the at least one poly(A) sequence is operably linked to at least one of the at least one nucleic acid sequences.

142. The composition, mixture, or method of any one of the above claims, 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, at least 90, at least 100, at least 110, or at least 120 consecutive A nucleotides.

143. The composition, mixture, or method of any one of the above claims, wherein the at least one poly(A) sequence is at least 80 consecutive A nucleotides.

144. The composition, mixture, or method of any one of the above claims, wherein the epitope-encoding nucleic acid sequence comprises an 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 a tumor, an infected cell, or an infectious disease organism, wherein the 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 a cell surface, optionally a tumor cell surface or an 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.

145. The composition of claim 74, 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 nucleotide sequencing data from a tumor, an infected cell, or an infectious disease organism, wherein the 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 a cell surface, optionally a tumor cell surface or an 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.

146. The composition of claim 144, wherein a number of the set of selected epitopes is 2-20.

147. The composition of any one of claims 144-146, 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 a cell surface, optionally a tumor cell surface or an 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.

148. The composition of any one of claims 144-147, wherein selecting the set of selected epitopes comprises selecting epitopes that have an increased likelihood of being presented on a cell surface, optionally a tumor cell surface or an infected cell surface, relative to unselected epitopes based on the presentation model.

149. The composition of any one of claims 144-148, wherein selecting the set of selected epitopes comprises selecting epitopes that have an increased likelihood of being capable of stimulating a tumor-specific or infectious disease organism-specific immune response in the subject relative to unselected epitopes based on the presentation model.

150. The composition of any one of claims 144-149, 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).

151. The composition of any one of claims 144-150, 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.

152. The composition of any one of claims 144-151, wherein selecting the set of selected epitopes comprises selecting epitopes that have a decreased likelihood of being capable of stimulating an autoimmune response to normal tissue in the subject relative to unselected epitopes based on the presentation model.

153. The composition of any one of claims 144-152, wherein exome or transcriptome nucleotide sequencing data is obtained by performing sequencing on a tumor cell or tissue, an infected cell, or an infectious disease organism.

154. The composition of claim 153, wherein the sequencing is next generation sequencing (NGS) or any massively parallel sequencing approach.

155. The composition, mixture, or method of any one of the above claims, wherein the composition for delivery of the self-amplifying expression system is administered as a priming vaccine.

156. The composition, mixture, or method of any one of the above claims, wherein the method further comprises administering a second composition, optionally wherein the second composition is a vaccine composition.

157. The method of claim 156, wherein the second composition is administered prior to the composition for delivery of the self-amplifying expression system.

158. The method of claim 156, wherein the second composition is administered subsequent to the administration of the composition for delivery of the self-amplifying expression system.

159. The method of any one of claims 156-158, wherein the second composition is the same as the composition for delivery of the self-amplifying expression system.

160. The method of any one of claims 156-158, wherein the second composition is different from the composition for delivery of the self-amplifying expression system.

161. The method of claim 160, wherein the second composition comprises the cassette of the self-amplifying expression system, optionally wherein the second composition comprises a chimpanzee adenovirus vector encoding the cassette of the self-amplifying expression system.
162. The method of any one of claims 156-161, wherein two or more second compositions are administered, optionally wherein the composition for delivery of the self-amplifying expression system is administered as a priming vaccine.
163. The method of any one of the above claims, wherein the composition for delivery of the self-amplifying expression system is administered intramuscularly (IM), intradermally (ID), subcutaneously (SC), intravitreal (IVT), intrathecal, or intravenously (IV).
164. The method of any one of the above claims, wherein the method further comprises administering an immune modulator, optionally 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, an anti-OX-40 antibody or an antigen-binding fragment thereof, or a cytokine, optionally wherein the cytokine is at least one of IL-2, IL-7, IL-12, IL-15, or IL-21 or variants thereof.
165. The method of any one of the above claims, wherein the method further comprises administering an adjuvant.

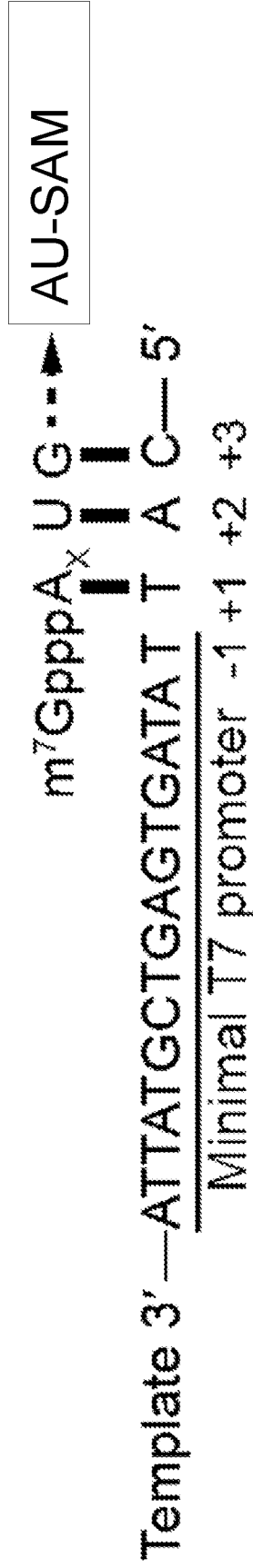
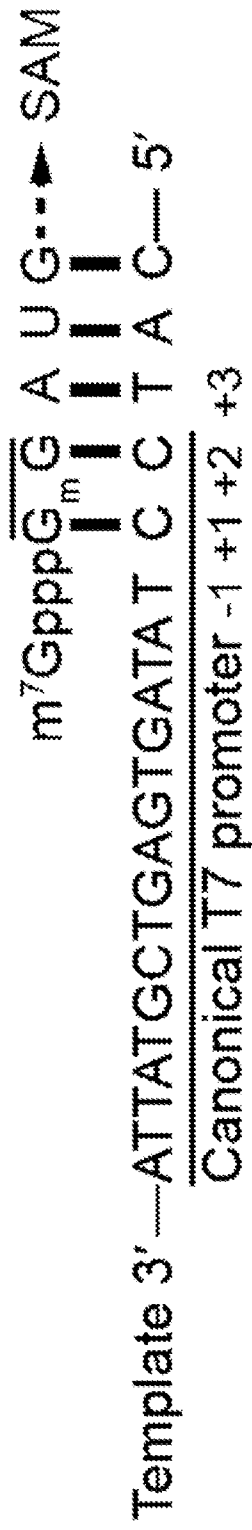


FIG. 1

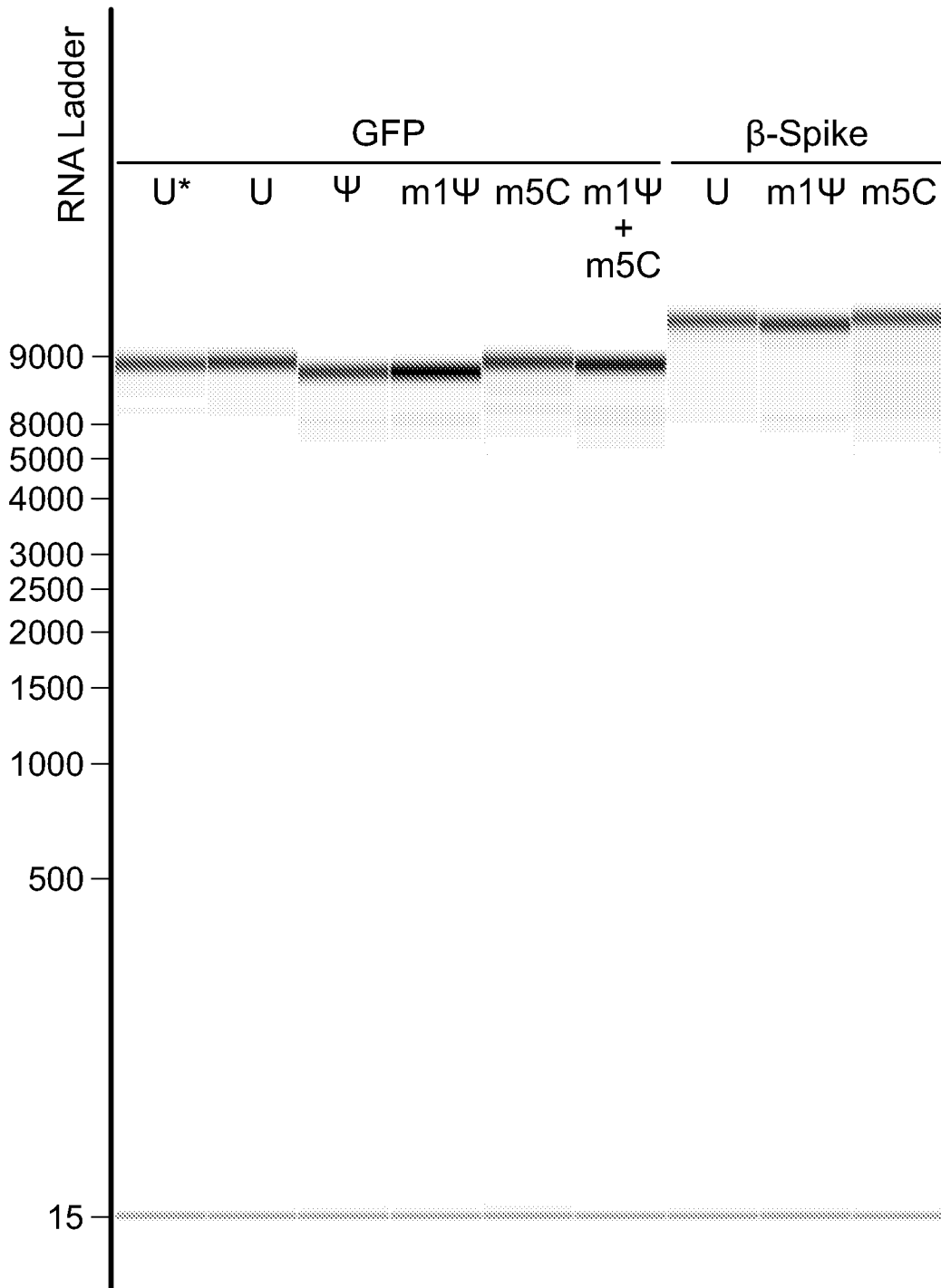


FIG. 3

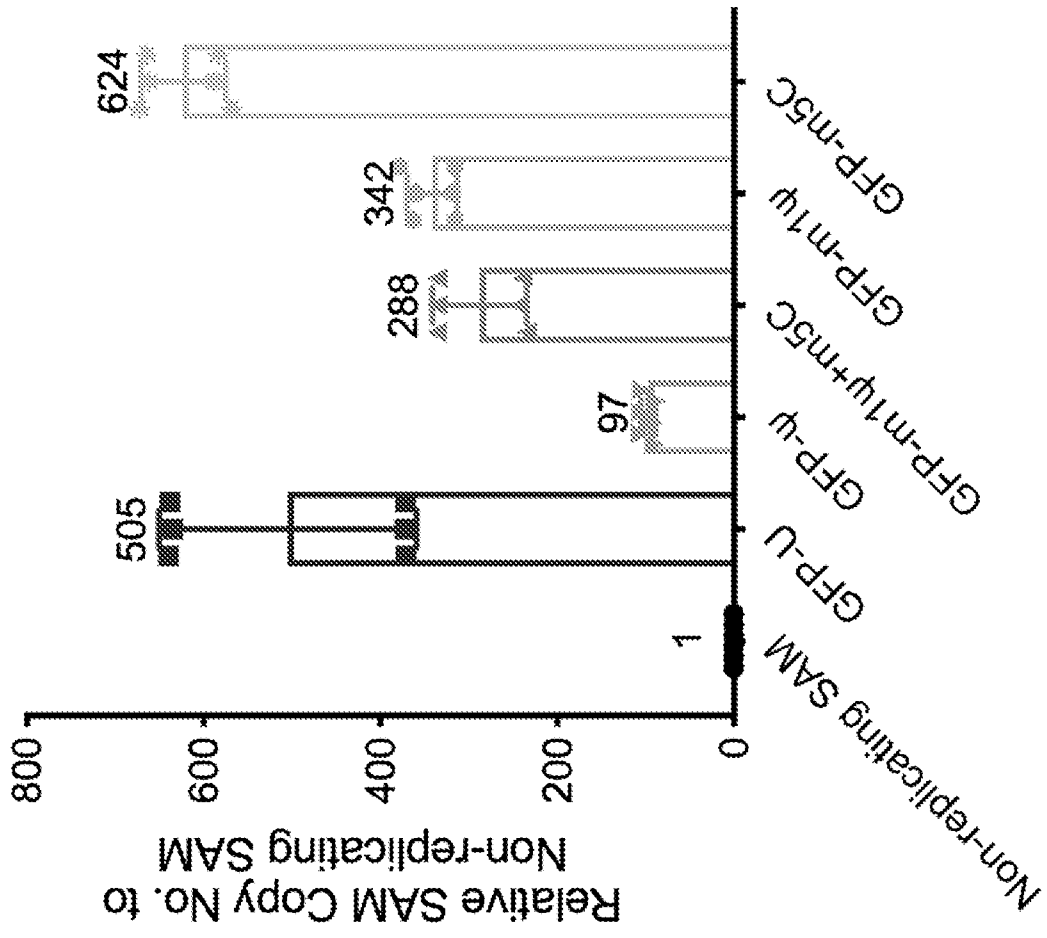


FIG. 4

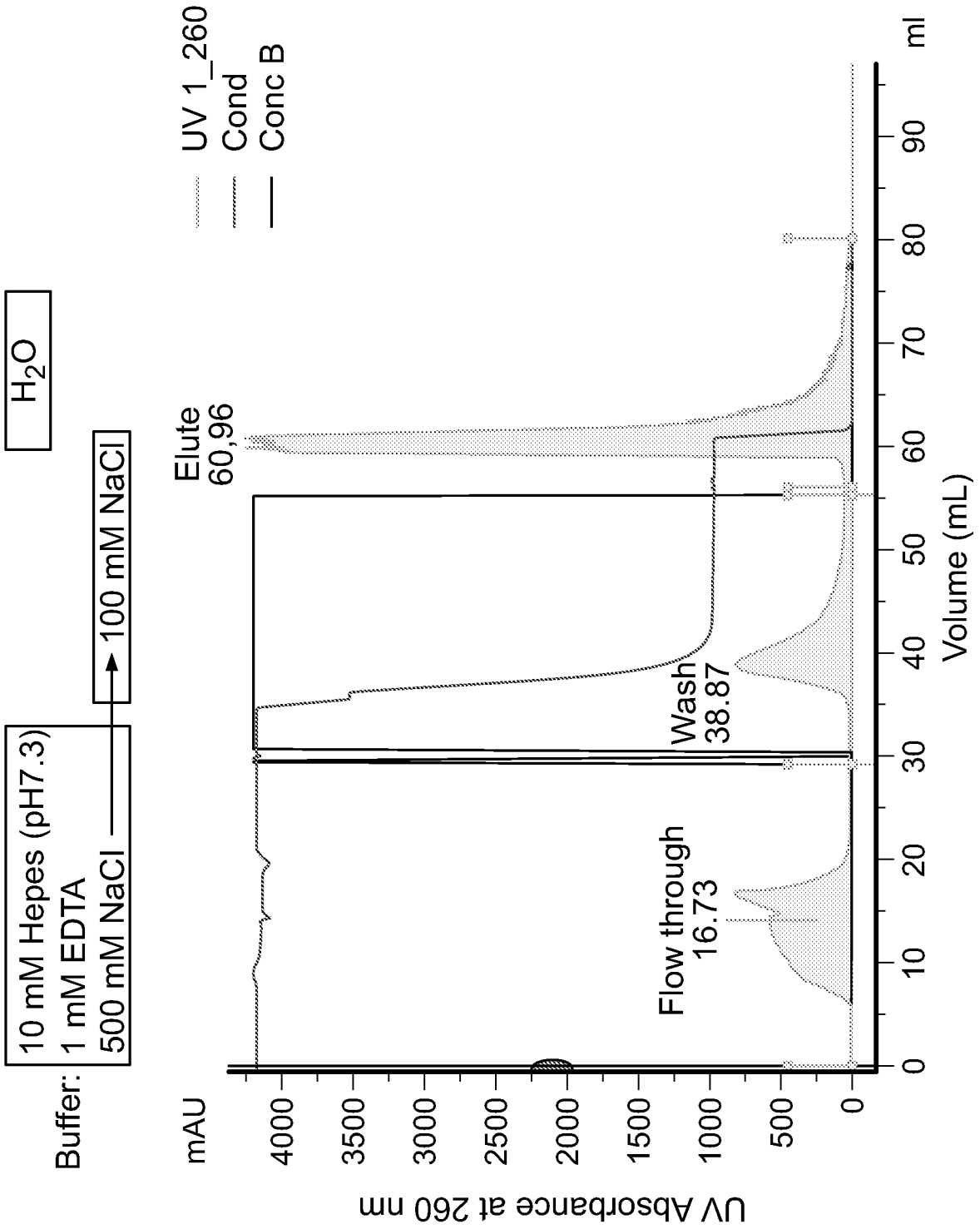


FIG. 5A

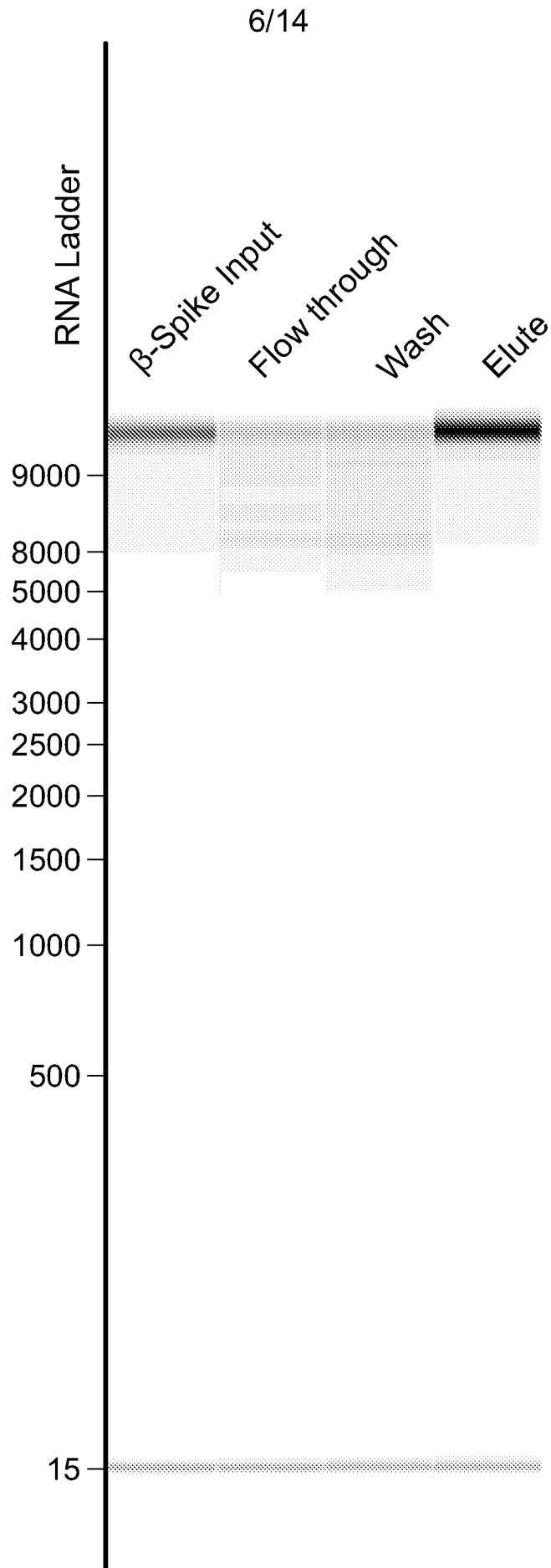


FIG. 5B

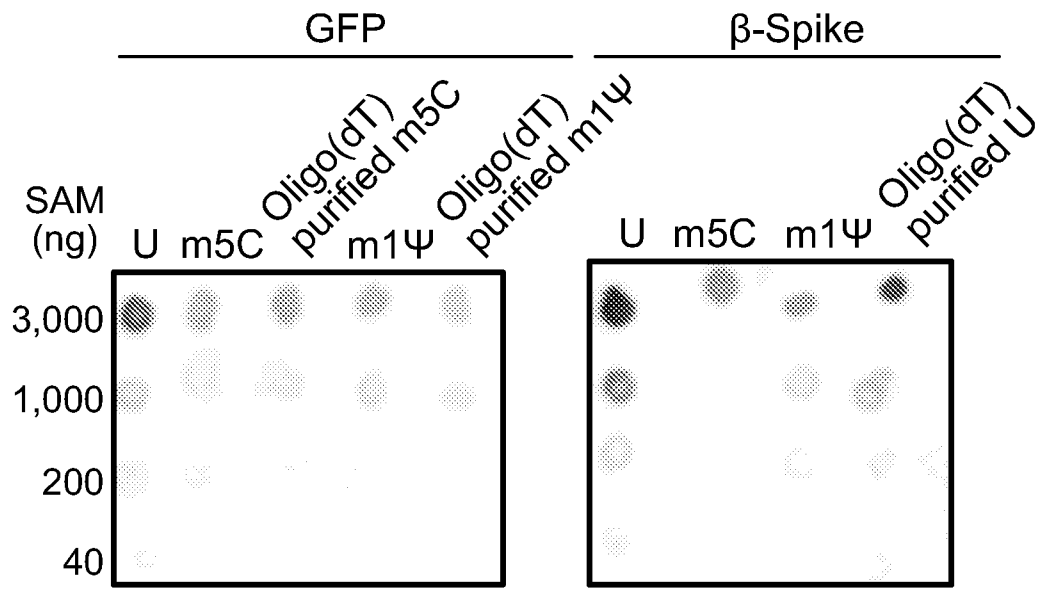


FIG. 6A

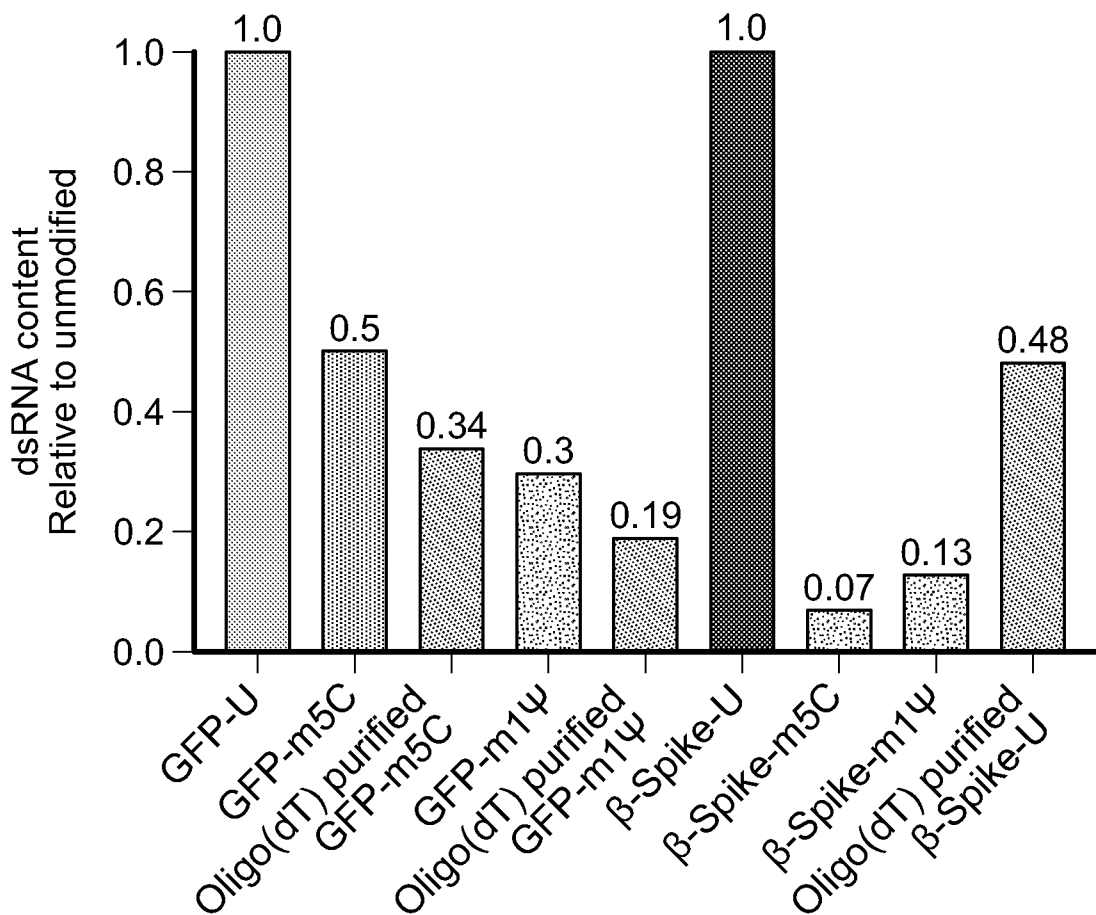


FIG. 6B

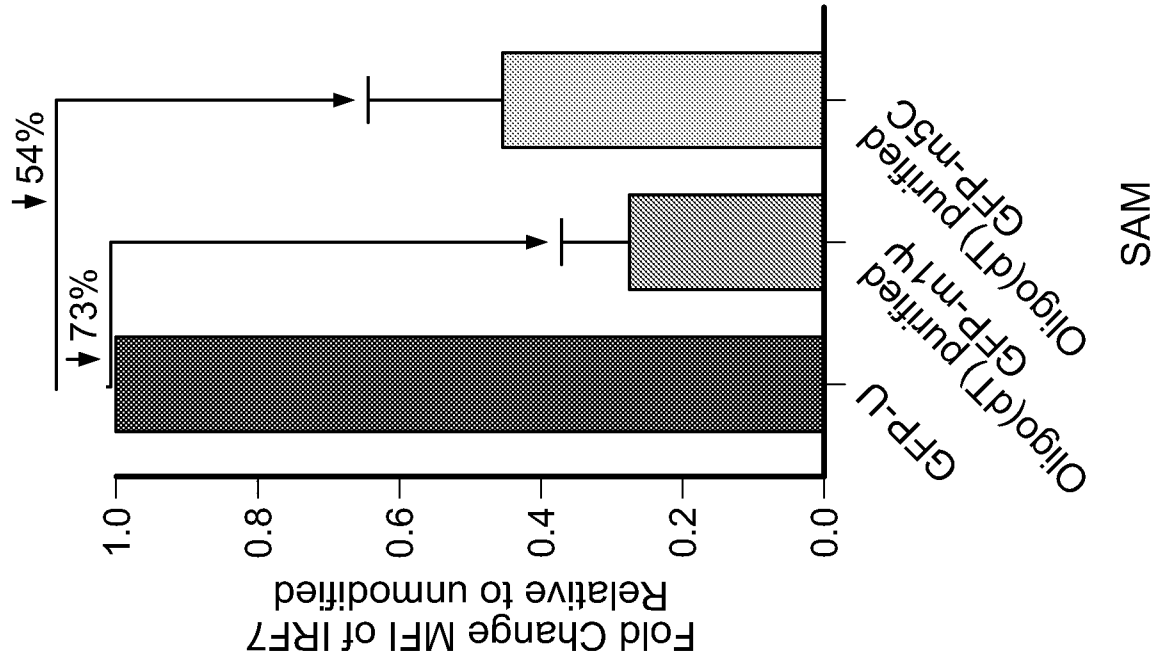


FIG. 7B

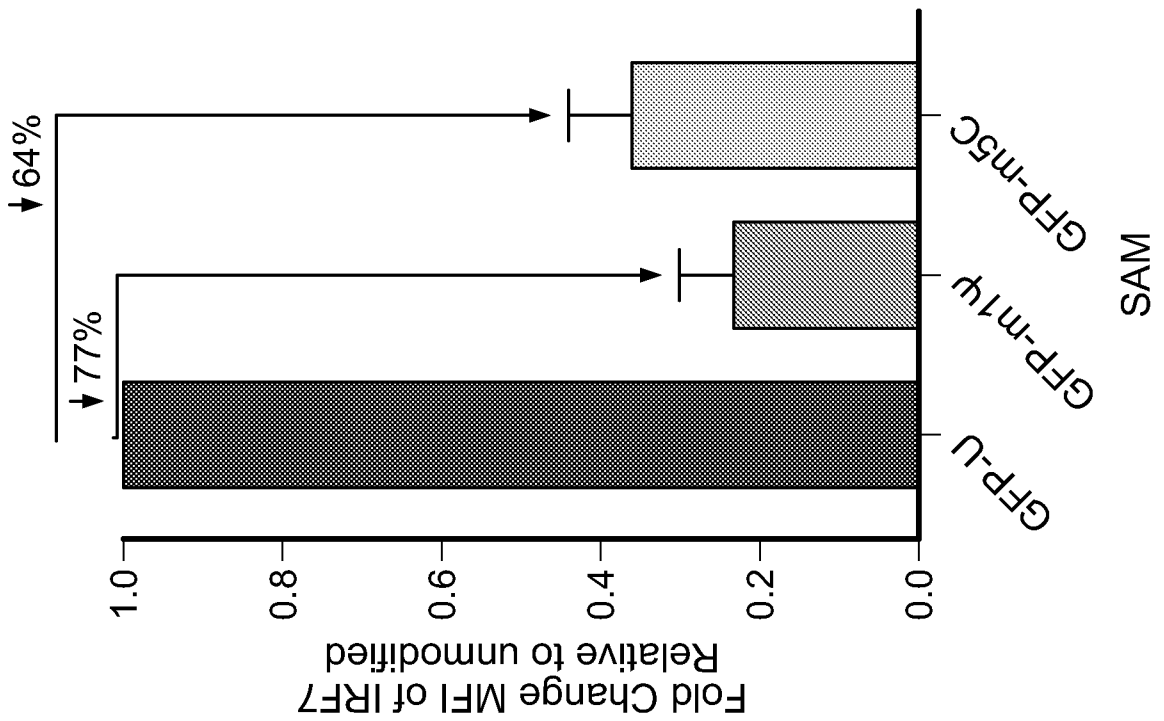


FIG. 7A

9/14

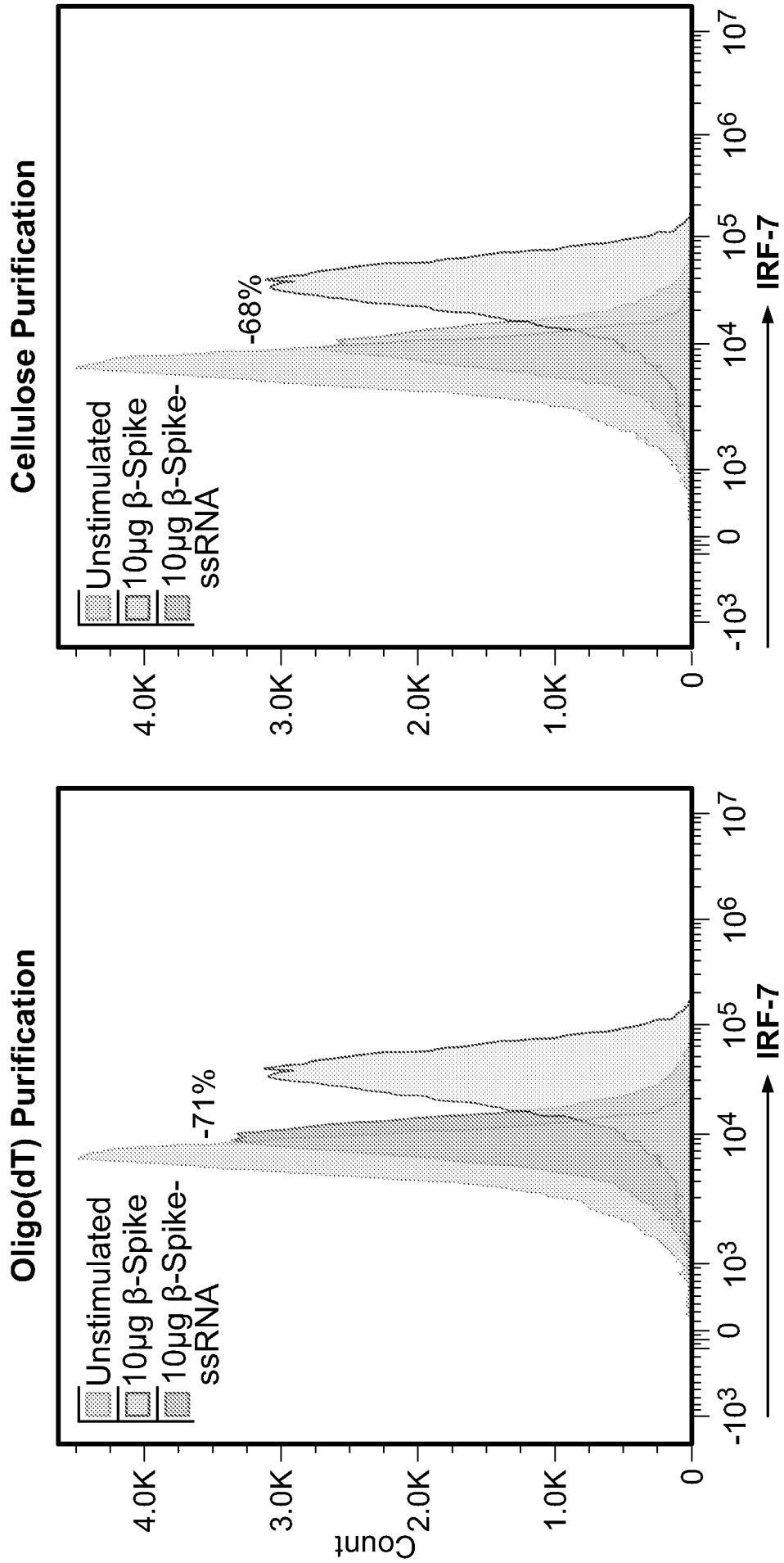


FIG. 7C

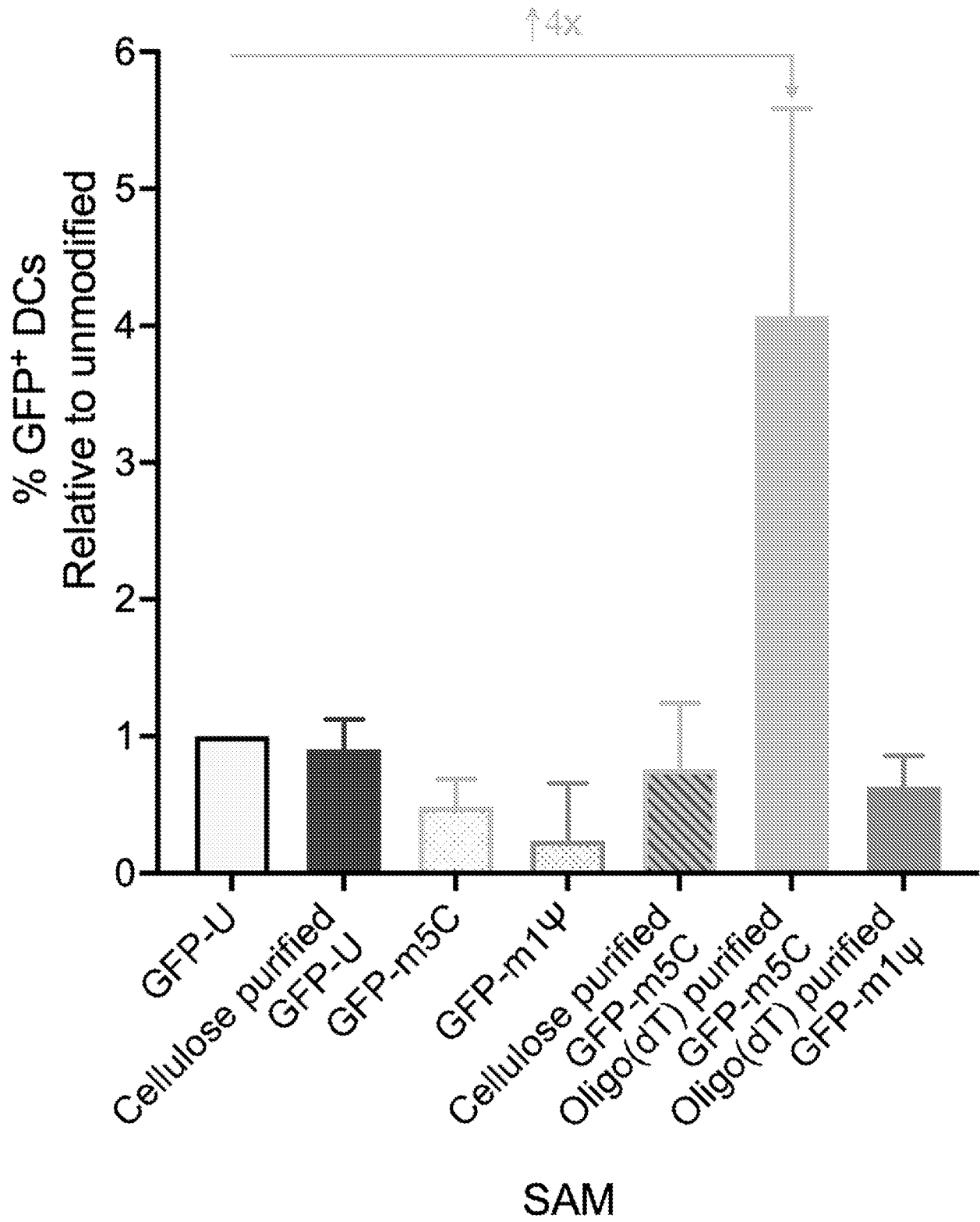


FIG. 8

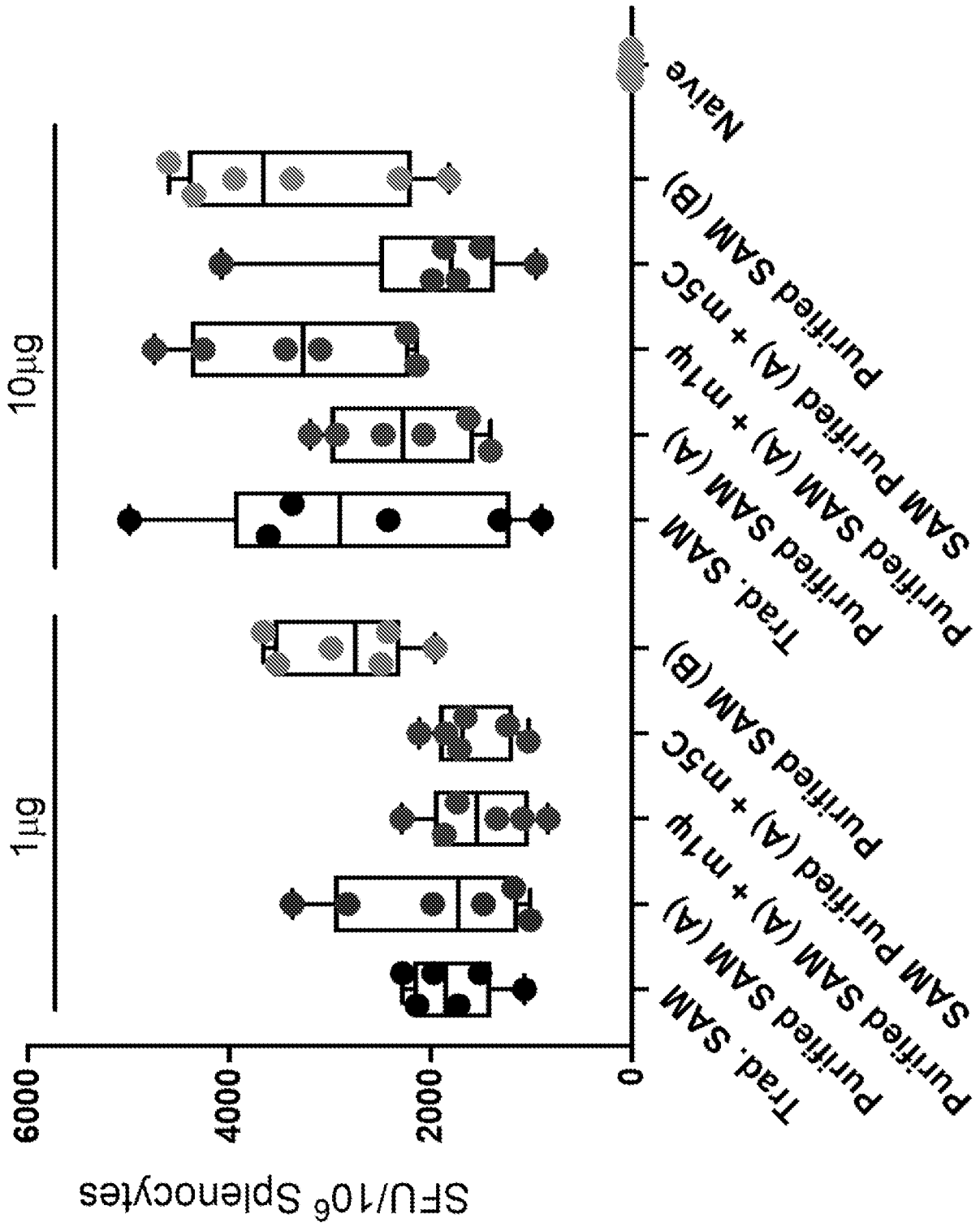


FIG. 9

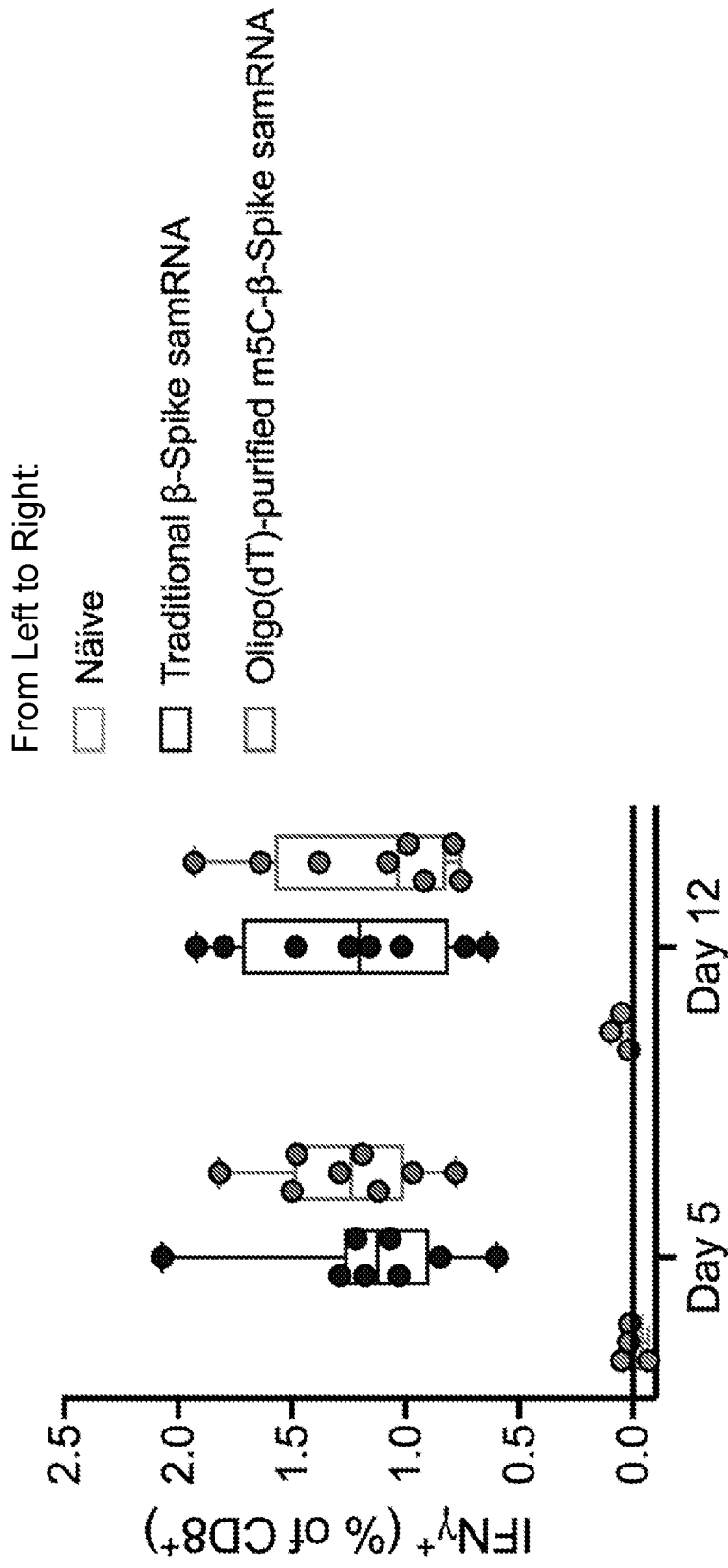


FIG. 10

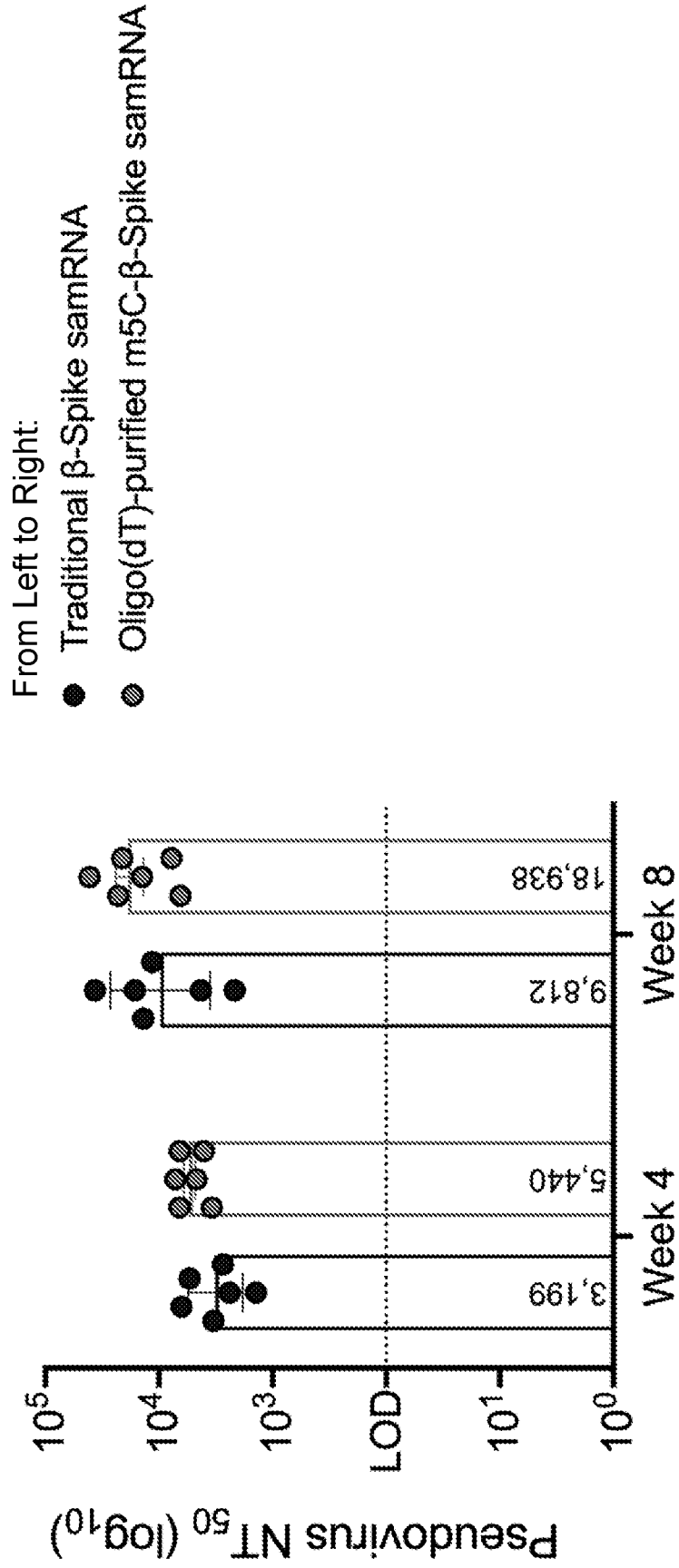


FIG. 11

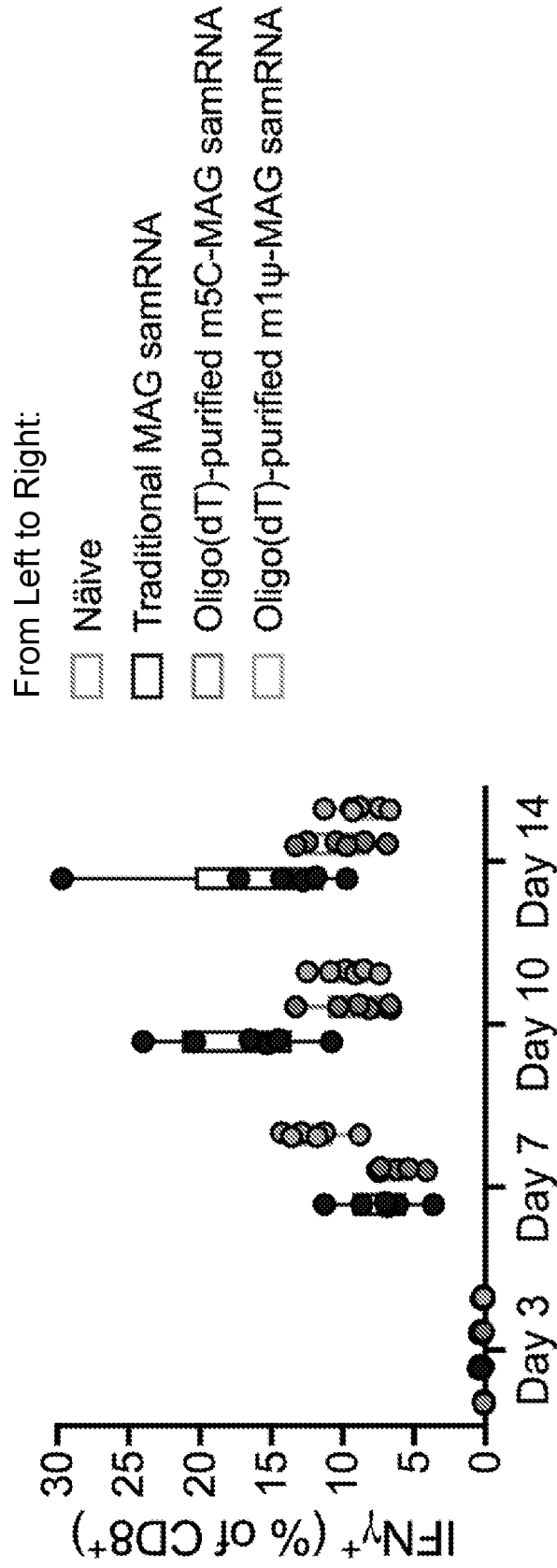


FIG. 12