



(51) International Patent Classification:

A61K 39/00 (2006.01) C07K 14/47 (2006.01)
A61P 35/00 (2006.01) C12N 15/861 (2006.01)

(21) International Application Number:

PCT/US2021/045106

(22) International Filing Date:

06 August 2021 (06.08.2021)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

63/062,268 06 August 2020 (06.08.2020) US

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

(72) Inventors: JOOSS, Karin; c/o Gritstone Bio, Inc., 5959 Horton Street, Suite 300, Emeryville, California 94608 (US). YELENSKY, Roman; c/o Gritstone Bio, Inc., 5959 Horton Street, Suite 300, Emeryville, California 94608 (US). SUN, James, Xin; c/o Gritstone Bio, Inc., 5959 Horton Street, Suite 300, Emeryville, California 94608 (US). RAPPAPORT, Amy, Rachel; c/o Gritstone Bio,

Inc., 5959 Horton Street, Suite 300, Emeryville, California 94608 (US). SCALLAN, Ciaran, Daniel; c/o Gritstone Bio, Inc., 5959 Horton Street, Suite 300, Emeryville, California 94608 (US). GITLIN, Leonid; c/o Gritstone Bio, Inc., 5959 Horton Street, Suite 300, Emeryville, California 94608 (US). PALMER, Christine, Denise; c/o Gritstone Bio, Inc., 5959 Horton Street, Suite 300, Emeryville, California 94608 (US). LANE, Monica; 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, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, IT, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW,

(54) Title: MULTIEPITOPE VACCINE CASSETTES

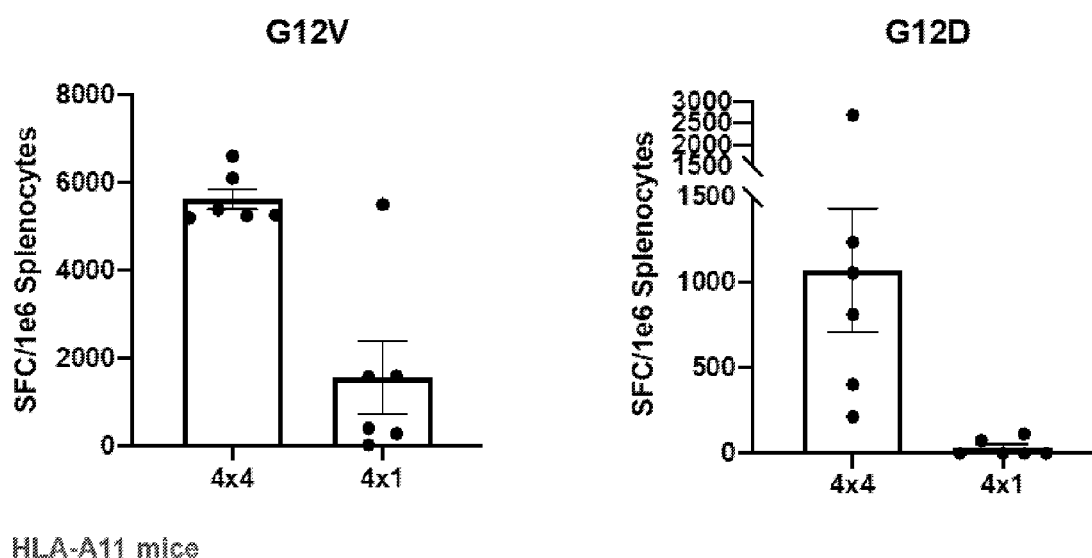


FIG. 3

(57) **Abstract:** Disclosed herein are compositions that include antigen-encoding nucleic acid sequences having multiple iterations of KRAS neoepitope-encoding sequences and/or lacking immunodominant epitopes. Also disclosed are nucleotides, cells, and methods associated with the compositions including their use as vaccines.



SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN,
TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

- (84) Designated States** (*unless otherwise indicated, for every kind of regional protection available*): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

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

Published:

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

(88) Date of publication of the international search report:

07 April 2022 (07.04.2022)

MULTIEPITOPE VACCINE CASSETTES

CROSS REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit of U.S. Provisional Application No. 63/062,268 filed August 6, 2020, which is hereby incorporated in its entirety by reference for all purposes.

SEQUENCE LISTING

[0002] The instant application contains a Sequence Listing which has been submitted electronically in ASCII format and is hereby incorporated by reference in its entirety. Said ASCII copy, created on August 6, 2020, is named GSO_095_sequencelisting.txt and is 425,984 bytes in size.

BACKGROUND

[0003] Therapeutic vaccines based on tumor-specific antigens hold great promise as a next-generation of personalized cancer immunotherapy.¹⁻³ For example, cancers with a high mutational burden, such as non-small cell lung cancer (NSCLC) and melanoma, are particularly attractive targets of such therapy given the relatively greater likelihood of neoantigen generation.^{4,5} Early evidence shows that neoantigen-based vaccination can elicit T-cell responses⁶ and that neoantigen targeted cell-therapy can cause tumor regression under certain circumstances in selected patients.⁷

[0004] One question for antigen vaccine design in both cancer and infectious disease settings is which of the many coding mutations present generate the “best” therapeutic antigens, e.g., antigens that can elicit immunity.

[0005] In addition to the challenges of current antigen prediction methods certain challenges also exist with the available vector systems that can be used for antigen delivery in humans, many of which are derived from humans. For example, many humans have pre-existing immunity to human viruses as a result of previous natural exposure, and this immunity can be a major obstacle to the use of recombinant human viruses for antigen delivery in vaccination strategies, such as in cancer treatment or vaccinations against infectious diseases. While some progress has been made in vaccinations strategies addressing the above problems, improvements are still needed, particularly for clinical applications, such as improved vaccine potency and efficacy.

SUMMARY

[0006] Disclosed herein is: an antigen-encoding cassette, or a polypeptide sequence encoded by the cassette, wherein the antigen-encoding cassette comprises at least one antigen-encoding nucleic acid sequence described, from 5' to 3', by the formula:

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

wherein E represents a nucleotide sequence comprising a 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, and at least one of the distinct epitope-encoding nucleic acid sequences comprising the at least two iterations encodes a distinct KRAS-associated MHC class I neoepitope.

[0007] In some aspects, the antigen-encoding cassette encodes at least 4 iterations of each of the amino acid sequences VVVGACGVGK (SEQ ID NO: 75), VVVGADGVGK (SEQ ID NO: 78), VVGAVGVGK (SEQ ID NO: 79), and ILDTAGHEEY (SEQ ID NO: 82). In some aspects, the KRAS-associated MHC class I neoepitope or the KRAS mutation comprises a KRAS G12C mutation, a KRAS G12V mutation, a KRAS G12D mutation, or a KRAS Q61H mutation. In some aspects, the KRAS-associated MHC class I neoepitope or the KRAS mutation comprises any one of the amino acid sequence shown in SEQ ID NOs: 75-82. In some aspects, the antigen-encoding cassette comprises each of the amino acid sequence shown in SEQ ID NOs: 75-82. In some aspects, the antigen-encoding cassette comprises two or more iterations of each of the amino acid sequence shown in SEQ ID NOs: 75-82. In some aspects, the antigen-encoding cassette comprises 4 iterations of each of the amino acid sequence shown in SEQ ID NOs: 75-82. In some aspects, the KRAS-associated MHC class I neoepitope or the KRAS mutation comprises the amino acid sequence shown in SEQ ID NO: 57, SEQ ID NO: 58, SEQ ID NO: 59, or SEQ ID NO: 60. In some aspects, the epitope-encoding nucleic acid sequences comprises two or more distinct epitope-encoding nucleic acid sequences independently encoding a distinct KRAS-associated MHC class I neoepitope or a distinct KRAS mutation. In some aspects, each of the epitope-encoding nucleic acid sequences independently encodes a distinct KRAS-associated MHC class I neoepitope or a distinct KRAS mutation. In some aspects, the epitope-encoding nucleic acid sequences comprises two or more distinct epitope-encoding nucleic acid sequences independently encoding a KRAS G12C mutation, a KRAS G12V mutation, a KRAS G12D

mutation, or a KRAS Q61H mutation. In some aspects, the epitope-encoding nucleic acid sequences independently encodes each of a KRAS G12C mutation, a KRAS G12V mutation, and a KRAS G12D mutation, and optionally a KRAS Q61H mutation. In some aspects, the antigen-encoding nucleic acid sequence encodes a peptide comprising the amino acid sequence shown in SEQ ID NO: 64 or SEQ ID NO: 65. In some aspects, the antigen-encoding nucleic acid sequence encodes a peptide comprising the amino acid sequence shown in SEQ ID NO: 65.

[0008] In some aspects, at least two of the distinct epitope-encoding nucleic acid sequences comprising the at least two iterations encode distinct KRAS-associated MHC class I neoepitopes. In some aspects, at least 3, at least 4, at least 5, at least 6, at least 7, or at least 8 of the distinct epitope-encoding nucleic acid sequences comprising the at least two iterations encode distinct KRAS-associated MHC class I neoepitopes. In some aspects, each of the distinct epitope-encoding nucleic acid sequences comprising the at least two iterations encode distinct KRAS-associated MHC class I neoepitopes. In some aspects, one or more of the nucleic acid sequences encoding the KRAS-associated MHC class I neoepitopes comprises at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, or at least 8 iterations. In some aspects, each of the nucleic acid sequences encoding the KRAS-associated MHC class I neoepitopes comprises at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, or at least 8 iterations. In some aspects, one or more of the nucleic acid sequences encoding the distinct KRAS-associated MHC class I neoepitopes comprises at least 4 iterations. In some aspects, each of the nucleic acid sequences encoding the distinct KRAS-associated MHC class I neoepitopes comprises at least 4 iterations. In some aspects, one or more of the distinct KRAS-associated MHC class I neoepitopes independently comprises a KRAS G12C mutation, a KRAS G12V mutation, a KRAS G12D mutation, or a KRAS Q61H mutation.

[0009] In some aspects, each E or E^N independently comprises a nucleotide sequence described, from 5' to 3', by the formula (L5_b-N_c-L3_d), wherein 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. In some aspects, each N encodes an 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 epitope, and wherein the 5' linker sequence encodes a peptide that is at least 2 amino acids in length, and L3 is a native 3' linker sequence that encodes a native C-terminal amino acid sequence of the epitope, and wherein the 3' linker sequence encodes a peptide that is at least 32 amino acids in length. In some aspects, the 5' and/or 3' linker sequence encodes a peptide that is at least 3 amino acids in length. In some aspects, the 5' and/or 3' linker sequence encodes a peptide that is at least 4 amino acids in length. In some aspects, the 5' and/or 3' linker sequence encodes a peptide that is at least 5

amino acids in length. In some aspects, the 5' and/or 3' linker sequence encodes a peptide that is at least 8 amino acids in length. In some aspects, the 5' and/or 3' linker sequence encodes a peptide that is at least 2-8 amino acids in length. In some aspects, the 5' and/or 3' linker sequence encodes a peptide that is at least 2-10 amino acids in length.

[0010] In some aspects, each E and E^N encodes an epitope at least 7 amino acids in length. In some aspects, each E and E^N encodes an epitope 7-15 amino acids in length. In some aspects, each E and E^N is a nucleotide sequence at least 21 nucleotides in length. In some aspects, each E and E^N is a nucleotide sequence 75 nucleotides in length.

[0011] Also disclosed herein is a composition for delivery of an antigen expression system, comprising: the antigen expression system, wherein the antigen expression system comprises one or more vectors, the one or more vectors comprising: (a) a vector backbone, wherein the backbone comprises: (i) at least one promoter nucleotide sequence, and (ii) optionally, at least one polyadenylation (poly(A)) sequence; and (b) a cassette, wherein the cassette comprises: (i) at least one antigen-encoding nucleic acid sequence, comprising: (I) an epitope-encoding nucleic acid sequence encoding a KRAS-associated MHC class I neoepitope, and wherein each of the epitope-encoding nucleic acid sequences comprises: (A) optionally, a 5' linker sequence, and (B) optionally, a 3' linker sequence; (ii) optionally, a second promoter nucleotide sequence operably linked to the antigen-encoding nucleic acid sequence; and (iii) optionally, at least one MHC class II epitope-encoding nucleic acid sequence; (iv) optionally, at least one nucleic acid sequence encoding a GPGPG amino acid linker sequence (SEQ ID NO:56); and (v) optionally, at least one second poly(A) sequence, wherein the second poly(A) sequence is a native poly(A) sequence or an exogenous poly(A) sequence to the vector backbone, wherein if the second promoter nucleotide sequence is absent, the antigen-encoding nucleic acid sequence is operably linked to the at least one promoter nucleotide sequence, and wherein the at least one antigen-encoding nucleic acid sequence comprises at least two iterations of the epitope-encoding nucleic acid sequence encoding the KRAS-associated MHC class I neoepitope.

[0012] Also disclosed herein is a composition for delivery of an antigen expression system, comprising: the antigen expression system, wherein the antigen expression system comprises one or more vectors, the one or more vectors comprising: (a) a vector backbone, wherein the backbone comprises: (i) at least one promoter nucleotide sequence, and (ii) at least one polyadenylation (poly(A)) sequence; and (b) a cassette, wherein the cassette comprises: (i) at least one antigen-encoding nucleic acid sequence, comprising: (I) at least 2, 3, 4, 5, 6, 7, 8, 9, or 10 distinct epitope-encoding nucleic acid sequences linearly linked to each other wherein at least one of the distinct epitope-encoding nucleic acid sequences encodes a KRAS-associated MHC

class I neoepitope, and wherein each of the epitope-encoding nucleic acid sequences comprises: (A) optionally, a 5' linker sequence, and (B) optionally, a 3' linker sequence; (ii) optionally, a second promoter nucleotide sequence operably linked to the antigen-encoding nucleic acid sequence; (iii) optionally, at least one MHC class II epitope-encoding nucleic acid sequence; (iv) optionally, at least one nucleic acid sequence encoding a GPGPG amino acid linker sequence (SEQ ID NO:56); and (v) optionally, at least one second poly(A) sequence, wherein the second poly(A) sequence is a native poly(A) sequence or an exogenous poly(A) sequence to the vector backbone, wherein if the second promoter nucleotide sequence is absent, the antigen-encoding nucleic acid sequence is operably linked to the at least one promoter nucleotide sequence, and wherein the at least one antigen-encoding nucleic acid sequence comprises at least two iterations of at least one of the distinct epitope-encoding nucleic acid sequences encoding the KRAS-associated MHC class I neoepitope.

[0013] In some aspects, the at least one antigen-encoding nucleic acid sequence comprises at least 3 distinct epitope-encoding nucleic acid sequences.

[0014] Also disclosed herein is a composition for delivery of an antigen expression system, comprising: the antigen expression system, wherein the antigen expression system comprises one or more vectors, the one or more vectors comprising: (a) a vector backbone, wherein the vector backbone comprises a chimpanzee adenovirus vector, optionally wherein the chimpanzee adenovirus vector is a ChAdV68 vector, or an alphavirus vector, optionally wherein the alphavirus vector is a Venezuelan equine encephalitis virus vector; and (b) a cassette, optionally wherein the cassette is integrated between a native promoter nucleotide sequence native to the vector backbone and a poly(A) sequence, optionally wherein the poly(A) sequence is native to the vector backbone, wherein the cassette comprises: (i) at least one antigen-encoding nucleic acid sequence, comprising: (I) an epitope-encoding nucleic acid sequence encoding a KRAS-associated MHC class I neoepitope, optionally comprising at least two distinct epitope-encoding nucleic acid sequences linearly linked to each other, each epitope-encoding nucleic acid sequence optionally comprising: (A) a MHC class I epitope encoding nucleic acid sequence, wherein the MHC class I epitope encoding nucleic acid sequence encodes a MHC class I epitope 7-15 amino acids in length, (B) a 5' linker sequence, wherein the 5' linker sequence encodes a native N-terminal amino acid sequence of the MHC class I epitope, and wherein the 5' linker sequence encodes a peptide that is at least 2 amino acids in length, (C) a 3' linker sequence, wherein the 3' linker sequence encodes a native C-terminal acid sequence of the MHC class I epitope, and wherein the 3' linker sequence encodes a peptide that is at least 2 amino acids in length, and wherein the cassette is operably linked to the native promoter nucleotide sequence, wherein each of the epitope-encoding nucleic acid sequences encodes a polypeptide that is

between 13 and 25 amino acids in length, and wherein each 3' end of each epitope-encoding nucleic acid sequence is linked to the 5' end of the following epitope-encoding nucleic acid sequence with the exception of the final epitope-encoding nucleic acid sequence in the cassette; and (ii) at least two MHC class II epitope-encoding nucleic acid sequences comprising: (I) a PADRE MHC class II sequence (SEQ ID NO:48), (II) a Tetanus toxoid MHC class II sequence (SEQ ID NO:46), (III) a first nucleic acid sequence encoding a GPGPG amino acid linker sequence linking the PADRE MHC class II sequence and the Tetanus toxoid MHC class II sequence, (IV) a second nucleic acid sequence encoding a GPGPG amino acid linker sequence linking the 5' end of the at least two MHC class II epitope-encoding nucleic acid sequences to the epitope-encoding nucleic acid sequences, (V) optionally, a third nucleic acid sequence encoding a GPGPG amino acid linker sequence at the 3' end of the at least two MHC class II epitope-encoding nucleic acid sequences; (iii) optionally, a second promoter nucleotide sequence operably linked to the antigen-encoding nucleic acid sequence; and wherein if the second promoter nucleotide sequence is absent, the antigen-encoding nucleic acid sequence is operably linked to the native promoter nucleotide sequence, and wherein the at least one antigen-encoding nucleic acid sequence comprises at least two iterations of the epitope-encoding nucleic acid sequence encoding the KRAS-associated MHC class I neoepitope.

[0015] In some aspects, an ordered sequence of each element of the cassette is described in the formula, from 5' to 3', comprising:

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

wherein, P comprises the second promoter nucleotide sequence, where a = 0 or 1, N comprises one of the distinct epitope-encoding nucleic acid sequences, 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 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.

[0016] In some aspects, for each X the corresponding N_c is a distinct epitope-encoding nucleic acid sequence, except for the N_c corresponding to the at least two iterations of the distinct epitope-encoding nucleic acid sequence. In some aspects, for each Y the corresponding U_f is a distinct MHC class II epitope-encoding nucleic acid sequence. In some aspects, a = 0, b = 1, d = 1, e = 1, g = 1, h = 1, X = 16, Y = 2, the at least one promoter nucleotide sequence is a single native promoter nucleotide sequence native to the vector backbone, the at least one

polyadenylation poly(A) sequence is a poly(A) sequence of at least 80 consecutive A nucleotides provided by the vector backbone, each N encodes an 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 epitope, and wherein the 5' linker sequence encodes a peptide that is at least 2 amino acids in length, L3 is a native 3' linker sequence that encodes a native C-terminal amino acid sequence of the epitope, and wherein the 3' linker sequence encodes a peptide that is at least 2 amino acids in length, U is each of a PADRE class II sequence and a Tetanus toxoid MHC class II sequence, the vector backbone comprises a chimpanzee adenovirus vector, optionally wherein the chimpanzee adenovirus vector is a ChAdV68 vector, or an alphavirus vector, optionally wherein the alphavirus vector is a Venezuelan equine encephalitis virus vector, optionally wherein the native promoter nucleotide sequence is a subgenomic (*e.g.*, 26S) promoter when the vector backbone comprises an alphavirus vector, and each of the MHC class II epitope-encoding nucleic acid sequences encodes a polypeptide that is between 13 and 25 amino acids in length.

[0017] In some aspects, the antigen-encoding cassette encodes at least 4 iterations of each of the amino acid sequences VVVGACGVGK (SEQ ID NO: 75), VVVGADGVGK (SEQ ID NO: 78), VVGAVGVGK (SEQ ID NO: 79), and ILDTAGHEEY (SEQ ID NO: 82). In some aspects, the KRAS-associated MHC class I neoepitope or the KRAS mutation comprises a KRAS G12C mutation, a KRAS G12V mutation, a KRAS G12D mutation, or a KRAS Q61H mutation. In some aspects, the KRAS-associated MHC class I neoepitope or the KRAS mutation comprises any one of the amino acid sequence shown in SEQ ID NOs: 75-82. In some aspects, the antigen-encoding cassette comprises each of the amino acid sequence shown in SEQ ID NOs: 75-82. In some aspects, the antigen-encoding cassette comprises two or more iterations of each of the amino acid sequence shown in SEQ ID NOs: 75-82. In some aspects, the antigen-encoding cassette comprises 4 iterations of each of the amino acid sequence shown in SEQ ID NOs: 75-82. In some aspects, the KRAS-associated MHC class I neoepitope or the KRAS mutation comprises the amino acid sequence shown in SEQ ID NO: 57, SEQ ID NO: 58, SEQ ID NO: 59, or SEQ ID NO: 60. In some aspects, the epitope-encoding nucleic acid sequences comprises two or more distinct epitope-encoding nucleic acid sequences independently encoding a distinct KRAS-associated MHC class I neoepitope or a distinct KRAS mutation. In some aspects, each of the epitope-encoding nucleic acid sequences independently encodes a distinct KRAS-associated MHC class I neoepitope or a distinct KRAS mutation. In some aspects, the epitope-encoding nucleic acid sequences comprises two or more distinct epitope-encoding nucleic acid sequences independently encoding a KRAS G12C mutation, a KRAS G12V mutation, a KRAS G12D mutation, or a KRAS Q61H mutation. In some aspects, the epitope-encoding nucleic acid sequences independently encodes each of a KRAS G12C mutation, a KRAS G12V mutation, and

a KRAS G12D mutation, and optionally a KRAS Q61H mutation. In some aspects, the antigen-encoding nucleic acid sequence encodes a peptide comprising the amino acid sequence shown in SEQ ID NO: 64 or SEQ ID NO: 65. In some aspects, the antigen-encoding nucleic acid sequence encodes a peptide comprising the amino acid sequence shown in SEQ ID NO: 65.

[0018] In some aspects, the at least two iterations is at least 3, at least 4, at least 5, at least 6, at least 7, or at least 8 iterations. In some aspects, the at least two iterations is at least 8 iterations. In some aspects, the at least two iterations is 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, or at least 20 iterations. In some aspects, the at least two iterations is between 2-3, between 2-4, between 2-5, between 2-6, between 2-7 iterations, or between 2-8 iterations. In some aspects, the at least two iterations is 7 iterations or less, 6 iterations or less, 5 iterations or less, 4 iterations or less, or 3 iterations or less.

[0019] In some aspects, the at least one antigen-encoding nucleic acid sequence comprises at least two iterations of at least two distinct epitope-encoding nucleic acid sequences. In some aspects, the at least one antigen-encoding nucleic acid sequence comprises at least two iterations of at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, or at least 10 distinct epitope-encoding nucleic acid sequences. In some aspects, the at least two iterations are separated by at least one separate distinct epitope-encoding nucleic acid sequence. In some aspects, the at least two iterations are separated by at least 2 separate distinct epitope-encoding nucleic acid sequences. In some aspects, the at least two iterations, inclusive of the optional 5' linker sequence and/or the optional 3' linker sequence, are separated by at least 75 nucleotides. In some aspects, the at least two iterations, inclusive of the optional 5' linker sequence and/or the optional 3' linker sequence, are separated by at least 150 nucleotides, at least 300 nucleotides, or at least 675 nucleotides. In some aspects, the at least two iterations, inclusive of the optional 5' linker sequence and/or the optional 3' linker sequence, are separated by at least 50 nucleotides, at least 100 nucleotides, at least 200 nucleotides, at least 250 nucleotides, at least 350 nucleotides, at least 400 nucleotides, at least 450 nucleotides, at least 500 nucleotides, at least 700 nucleotides, at least 750 nucleotides, at least 800 nucleotides, at least 900 nucleotides, or at least 1000 nucleotides. In some aspects, the at least two iterations, inclusive of the optional 5' linker sequence and/or the optional 3' linker sequence, are separated by at least 10 nucleotides, at least 15 nucleotides, at least 20 nucleotides, at least 25 nucleotides, at least 30 nucleotides, at least 35 nucleotides, at least 40 nucleotides, at least 45 nucleotides, at least 50 nucleotides, at least 55 nucleotides, at least 60 nucleotides, at least 65 nucleotides, or at least 70 nucleotides.

[0020] In some aspects, the at least one antigen-encoding nucleic acid sequence is described, from 5' to 3', by the formula:

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

wherein, E represents a nucleotide sequence comprising at least one of the 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.

[0021] In some aspects, the antigen-encoding cassette encodes at least 4 iterations of each of the amino acid sequences VVVGACGVGK (SEQ ID NO: 75), VVVGADGVGK (SEQ ID NO: 78), VVGAVGVGK (SEQ ID NO: 79), and ILDTAGHEEY (SEQ ID NO: 82).

[0022] In some aspects, the distinct epitope-encoding nucleic acid sequences comprises at least two distinct epitope-encoding nucleic acid sequences each encoding distinct KRAS-associated MHC class I neoepitopes. In some aspects, the distinct epitope-encoding nucleic acid sequences comprises at least 3, at least 4, at least 5, at least 6, at least 7, or at least 8 distinct epitope-encoding nucleic acid sequences each encoding distinct KRAS-associated MHC class I neoepitopes. In some aspects, each of the epitope-encoding nucleic acid sequences of the at least one antigen-encoding nucleic acid sequence encodes a distinct KRAS-associated MHC class I neoepitope. In some aspects, one or more of the nucleic acid sequences encoding the distinct KRAS-associated MHC class I neoepitopes comprises at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, or at least 8 iterations. In some aspects, each of the nucleic acid sequences encoding the distinct KRAS-associated MHC class I neoepitopes comprises at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, or at least 8 iterations. In some aspects, one or more of the nucleic acid sequences encoding the distinct KRAS-associated MHC class I neoepitopes comprises at least 4 iterations. In some aspects, each of the nucleic acid sequences encoding the distinct KRAS-associated MHC class I neoepitopes comprises at least 4 iterations. In some aspects, one or more of the distinct KRAS-associated MHC class I neoepitopes independently comprises a KRAS G12C mutation, a KRAS G12V mutation, a KRAS G12D mutation, or a KRAS Q61H mutation.

[0023] In some aspects, the at least two iterations comprises a number of iterations, or z comprises a number, sufficient to stimulate a greater immune response relative to an antigen-encoding nucleic acid sequence comprising a single iteration of the epitope-encoding nucleic acid sequence. In some aspects, the at least two iterations comprises a number of iterations, or z

comprises a number, sufficient to stimulate an immune response, and a single iteration of the epitope-encoding nucleic acid sequence is insufficient to stimulate the immune response or insufficient to stimulate a detectable immune response. In some aspects, the immune response is an expansion of epitope-specific T cells following *in vivo* immunization with the composition for delivery of the antigen expression system. In some aspects, the immune response is increased activation of epitope-specific T cells and/or increased epitope-specific killing by epitope-specific T cells following *in vivo* immunization with the composition for delivery of the antigen expression system.

[0024] Also provided for herein is a composition for delivery of an antigen expression system, comprising: the antigen expression system, wherein the antigen expression system comprises one or more vectors, the one or more vectors comprising: (a) a vector backbone, wherein the backbone comprises: (i) at least one promoter nucleotide sequence, and (ii) optionally, at least one polyadenylation (poly(A)) sequence; and (b) a cassette, wherein the cassette comprises: (i) at least one antigen-encoding nucleic acid sequence, comprising: (I) at least two distinct epitope-encoding nucleic acid sequences, 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, optionally wherein the at least one alteration is a KRAS mutation, 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, and wherein each of the epitope-encoding nucleic acid sequences comprises: (A) optionally, a 5' linker sequence, and (B) optionally, a 3' linker sequence; (ii) optionally, a second promoter nucleotide sequence operably linked to the antigen-encoding nucleic acid sequence; and (iii) optionally, at least one MHC class II epitope-encoding nucleic acid sequence; (iv) optionally, at least one nucleic acid sequence encoding a GPGPG amino acid linker sequence (SEQ ID NO:56); and (v) optionally, at least one second poly(A) sequence, wherein the second poly(A) sequence is a native poly(A) sequence or an exogenous poly(A) sequence to the vector backbone, wherein if the second promoter nucleotide sequence is absent, the antigen-encoding nucleic acid sequence is operably linked to the at least one promoter nucleotide sequence, and wherein the cassette does not encode an immunodominant MHC class I epitope that: (1) stimulates a 5-fold or greater immune response when administered in a vaccine composition to a subject relative to another MHC class I epitope encoded in the cassette and capable of stimulating an immune response in the subject, and/or (2) reduces an immune response to another MHC class I epitope encoded in the cassette when administered in a vaccine composition to a subject relative to an immune response when the other MHC class I epitope is

administered in the absence of the immunodominant MHC class I epitope, optionally wherein the immune response is reduced to below a limit of detection and/or wherein the immune response is not a therapeutically effective response.

[0025] Also provided for herein is a composition for delivery of an antigen expression system, comprising: the antigen expression system, wherein the antigen expression system comprises one or more vectors, the one or more vectors comprising: (a) a vector backbone, wherein the backbone comprises: (i) at least one promoter nucleotide sequence, and (ii) optionally, at least one polyadenylation (poly(A)) sequence; and (b) a cassette, wherein the cassette comprises: (i) at least one antigen-encoding nucleic acid sequence, comprising: (I) at least 2, 3, 4, 5, 6, 7, 8, 9, or 10 distinct epitope-encoding nucleic acid sequences linearly linked to each other, 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, optionally wherein the at least one alteration is a KRAS mutation, 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, and wherein each of the epitope-encoding nucleic acid sequences comprises: (A) optionally, a 5' linker sequence, and (B) optionally, a 3' linker sequence; (ii) optionally, a second promoter nucleotide sequence operably linked to the antigen-encoding nucleic acid sequence; and (iii) optionally, at least one MHC class II epitope-encoding nucleic acid sequence; (iv) optionally, at least one nucleic acid sequence encoding a GP GPG amino acid linker sequence (SEQ ID NO:56); and (v) optionally, at least one second poly(A) sequence, wherein the second poly(A) sequence is a native poly(A) sequence or an exogenous poly(A) sequence to the vector backbone, and wherein if the second promoter nucleotide sequence is absent, the antigen-encoding nucleic acid sequence is operably linked to the at least one promoter nucleotide sequence, and wherein the cassette does not encode an immunodominant MHC class I epitope that: (1) stimulates a 5-fold or greater immune response when administered in a vaccine composition to a subject relative to another MHC class I epitope encoded in the cassette and capable of stimulating an immune response in the subject, and/or (2) reduces an immune response to another MHC class I epitope encoded in the cassette when administered in a vaccine composition to a subject relative to an immune response when the other MHC class I epitope is administered in the absence of the immunodominant MHC class I epitope, optionally wherein the immune response is reduced to below a limit of detection and/or wherein the immune response is not a therapeutically effective response.

[0026] In some aspects, at least one of the distinct epitope-encoding nucleic acid sequences encodes a KRAS-associated MHC class I neoepitope.

[0027] Also provided for herein is a composition for delivery of an antigen expression system, comprising: the antigen expression system, wherein the antigen expression system comprises one or more vectors, the one or more vectors comprising: (a) a vector backbone, wherein the vector backbone comprises a chimpanzee adenovirus vector, optionally wherein the chimpanzee adenovirus vector is a ChAdV68 vector, or an alphavirus vector, optionally wherein the alphavirus vector is a Venezuelan equine encephalitis virus vector; and (b) a cassette, optionally wherein the cassette is integrated between a native promoter nucleotide sequence native to the vector backbone and a poly(A) sequence, optionally wherein the poly(A) sequence is native to the vector backbone, wherein the cassette comprises: (i) at least one antigen-encoding nucleic acid sequence, comprising: (I) at least one epitope-encoding nucleic acid sequence, optionally comprising at least two distinct epitope-encoding nucleic acid sequences linearly linked to each other, each epitope-encoding nucleic acid sequence optionally comprising: (A) a MHC class I epitope encoding nucleic acid sequence, wherein the MHC class I epitope encoding nucleic acid sequence encodes a MHC class I epitope 7-15 amino acids in length, (B) a 5' linker sequence, wherein the 5' linker sequence encodes a native N-terminal amino acid sequence of the MHC class I epitope, and wherein the 5' linker sequence encodes a peptide that is at least 2 amino acids in length, (C) a 3' linker sequence, wherein the 3' linker sequence encodes a native C-terminal acid sequence of the MHC class I epitope, and wherein the 3' linker sequence encodes a peptide that is at least 2 amino acids in length, and wherein the cassette is operably linked to the native promoter nucleotide sequence, wherein each of the epitope-encoding nucleic acid sequences encodes a polypeptide that is between 13 and 25 amino acids in length, and wherein each 3' end of each epitope-encoding nucleic acid sequence is linked to the 5' end of the following epitope-encoding nucleic acid sequence with the exception of the final epitope-encoding nucleic acid sequence in the cassette; and (ii) at least two MHC class II epitope-encoding nucleic acid sequences comprising: (I) a PADRE MHC class II sequence (SEQ ID NO:48), (II) a Tetanus toxoid MHC class II sequence (SEQ ID NO:46), (III) a first nucleic acid sequence encoding a GPGPG amino acid linker sequence linking the PADRE MHC class II sequence and the Tetanus toxoid MHC class II sequence, (IV) a second nucleic acid sequence encoding a GPGPG amino acid linker sequence linking the 5' end of the at least two MHC class II epitope-encoding nucleic acid sequences to the epitope-encoding nucleic acid sequences, (V) optionally, a third nucleic acid sequence encoding a GPGPG amino acid linker sequence at the 3' end of the at least two MHC class II epitope-encoding nucleic acid sequences, and (iii) optionally, a second promoter nucleotide sequence operably linked to the antigen-encoding nucleic acid sequence; and wherein if the second promoter nucleotide sequence is absent, the antigen-encoding nucleic acid sequence is operably linked to the native promoter nucleotide

sequence, and wherein the cassette does not encode an immunodominant MHC class I epitope that: (1) stimulates a 5-fold or greater immune response when administered in a vaccine composition to a subject relative to another MHC class I epitope encoded in the cassette and capable of stimulating an immune response in the subject, and/or (2) reduces an immune response to another MHC class I epitope encoded in the cassette when administered in a vaccine composition to a subject relative to an immune response when the other MHC class I epitope is administered in the absence of the immunodominant MHC class I epitope, optionally wherein the immune response is reduced to below a limit of detection and/or wherein the immune response is not a therapeutically effective response.

[0028] In some aspects, the immunodominant MHC class I epitope stimulates a 10-fold, 20-fold, 30-fold, 40-fold, 50-fold, 100-fold, 200-fold, 300-fold, 400-fold, 500-fold, 1000-fold, 2000-fold, 3000-fold, 4000-fold, 5000-fold, or 10,000-fold or greater immune response when administered in a vaccine composition to a subject relative to another MHC class I epitope encoded in the cassette and capable of stimulating an immune response in the subject. In some aspects, the immunodominant MHC class I epitope reduces the immune response of the other MHC class I epitope to below a limit of detection and/or does not stimulate a therapeutically effective response. In some aspects, the subject expresses at least one HLA allele known or predicted to present both the immunodominant MHC class I epitope and the other MHC class I epitope encoded in the cassette.

[0029] In some aspects, one or more of the epitope-encoding nucleic acid sequences are derived from a tumor, an infection, or an infected cell of a subject. In some aspects, each of the epitope-encoding nucleic acid sequences are derived from a tumor, an infection, or an infected cell of a subject. In some aspects, one or more of the epitope-encoding nucleic acid sequences are not derived from a tumor, an infection, or an infected cell of a subject. In some aspects, each of the epitope-encoding nucleic acid sequences are not derived from a tumor, an infection, or an infected cell of a subject.

[0030] In some aspects, the epitope-encoding nucleic acid sequence encodes an epitope known or suspected to be presented by MHC class I on a surface of a cell, optionally wherein the surface of the cell is a tumor cell surface or an infected cell surface, and optionally wherein the cell is a subject's cell. In some aspects, the cell is a tumor cell 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, brain cancer, B-cell lymphoma, acute myelogenous leukemia, chronic myelogenous leukemia, chronic lymphocytic leukemia, T cell lymphocytic leukemia, non-small cell lung cancer, and small cell lung cancer, or wherein the cell is an infected cell selected from the group consisting of: a

pathogen infected cell, a virally infected cell, a bacterially infected cell, a fungally infected cell, and a parasitically infected cell. In some aspects, the virally infected cell is selected from the group consisting of: an HIV infected cell, a Severe acute respiratory syndrome-related coronavirus (SARS) infected cell, a severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infected cell, a Ebola infected cell, a Hepatitis B virus (HBV) infected cell, an influenza infected cell, an orthomyxoviridae family virus infected cell, a Human papillomavirus (HPV) infected cell, a Cytomegalovirus (CMV) infected cell, a Chikungunya virus infected cell, a Respiratory syncytial virus (RSV) infected cell, a Dengue virus infected cell, and a Hepatitis C virus (HCV) infected cell.

[0031] In some aspects, the composition further comprises a nanoparticulate delivery vehicle. In some aspects, the nanoparticulate delivery vehicle is a lipid nanoparticle (LNP). In some aspects, the LNP comprises ionizable amino lipids. In some aspects, the ionizable amino lipids comprise MC3-like (dilinoleylmethyl-4-dimethylaminobutyrate) molecules. In some aspects, the nanoparticulate delivery vehicle encapsulates the antigen expression system.

[0032] 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 second promoter is absent and the at least one promoter nucleotide sequence is operably linked to the antigen-encoding nucleic acid sequence.

[0033] In some aspects, the one or more vectors comprise one or more +-stranded RNA vectors. In some aspects, the one or more +-stranded RNA vectors comprise a 5' 7-methylguanosine (m7g) cap. In some aspects, the one or more +-stranded RNA vectors are produced by *in vitro* transcription. In some aspects, the one or more vectors are self-replicating within a mammalian cell. In some aspects, the backbone comprises at least one nucleotide sequence of an Aura virus, a Fort Morgan virus, a Venezuelan equine encephalitis virus, a Ross River virus, a Semliki Forest virus, a Sindbis virus, or a Mayaro virus. In some aspects, the backbone comprises at least one nucleotide sequence of a Venezuelan equine encephalitis virus. In some aspects, the backbone comprises at least sequences for nonstructural protein-mediated amplification, a 26S promoter sequence, a poly(A) sequence, a nonstructural protein 1 (nsP1) gene, a nsP2 gene, a nsP3 gene, and a nsP4 gene encoded by the nucleotide sequence of the Aura virus, the Fort Morgan virus, the Venezuelan equine encephalitis virus, the Ross River virus, the Semliki Forest virus, the Sindbis virus, or the Mayaro virus. In some aspects, the backbone comprises at least sequences for nonstructural protein-mediated amplification, a 26S promoter sequence, and a poly(A) sequence encoded by the nucleotide sequence of the Aura virus, the Fort Morgan virus, the Venezuelan equine encephalitis virus, the Ross River virus, the Semliki Forest virus, the Sindbis virus, or the Mayaro virus. In some aspects, sequences for nonstructural

protein-mediated amplification are selected from the group consisting of: an alphavirus 5' UTR, a 51-nt CSE, a 24-nt CSE, a 26S subgenomic promoter sequence, a 19-nt CSE, an alphavirus 3' UTR, or combinations thereof. In some aspects, the backbone does not encode structural virion proteins capsid, E2 and E1. In some aspects, the cassette is inserted in place of structural virion proteins within the nucleotide sequence of the Aura virus, the Fort Morgan virus, the Venezuelan equine encephalitis virus, the Ross River virus, the Semliki Forest virus, the Sindbis virus, or the Mayaro virus.

[0034] 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 backbone comprises the sequence set forth in SEQ ID NO:6 or SEQ ID NO:7. In some aspects, the cassette is inserted at position 7544 to replace the deletion between base pairs 7544 and 11175 as set forth in the sequence of SEQ ID NO:3 or SEQ ID NO:5.

[0035] 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 antigen-encoding nucleic acid sequence, wherein the nsP1-4 genes and the at least one antigen-encoding nucleic acid sequence are in separate open reading frames.

[0036] In some aspects, the backbone comprises at least one nucleotide sequence of a chimpanzee adenovirus vector. In some aspects, the chimpanzee adenovirus vector is a ChAdV68 vector. In some aspects, the ChAdV68 vector comprises a ChAdV68 vector backbone comprising the sequence set forth in SEQ ID NO:1. In some aspects, the ChAdV68 vector comprises a ChAdV68 vector backbone comprising the sequence set forth in SEQ ID NO:1, except that the sequence is fully deleted or functionally deleted in at least one gene selected from the group consisting of the chimpanzee adenovirus E1A, E1B, E2A, E2B, E3, E4, L1, L2, L3, L4, and L5 genes of the sequence set forth in SEQ ID NO:1, optionally wherein the sequence is fully deleted or functionally deleted in: (1) E1A and E1B; (2) E1A, E1B, and E3; or (3) E1A, E1B, E3, and E4 of the sequence set forth in SEQ ID NO:1. In some aspects, the ChAdV68 vector comprises a ChAdV68 vector backbone comprising a gene or regulatory sequence obtained from the sequence of SEQ ID NO:1, optionally wherein the gene is selected from the group consisting of the chimpanzee adenovirus inverted terminal repeat (ITR), E1A, E1B, E2A, E2B, E3, E4, L1, L2, L3, L4, and L5 genes of the sequence set forth in SEQ ID NO:1. In some aspects, the ChAdV68 vector comprises a ChAdV68 vector backbone comprising a partially deleted E4 gene comprising a deleted or partially-deleted E4orf2 region and a deleted or partially-deleted E4orf3 region, and optionally a deleted or partially-deleted E4orf4 region. In

some aspects, the ChAdV68 vector comprises a ChAdV68 vector backbone comprising at least nucleotides 2 to 36,518 of the sequence set forth in SEQ ID NO:1 and further comprising: (1) an E1 deletion of at least nucleotides 577 to 3403 of the sequence shown in SEQ ID NO:1, (2) an E3 deletion of at least nucleotides 27,125 to 31,825 of the sequence shown in SEQ ID NO:1, and (3) an E4 deletion of at least nucleotides 34,916 to 35,642 of the sequence shown in SEQ ID NO:1; optionally wherein the antigen cassette is inserted within the E1 deletion. In some aspects, the ChAdV68 vector comprises a ChAdV68 vector backbone comprising the sequence set forth in SEQ ID NO:68, optionally wherein the antigen cassette is inserted within the E1 deletion. In some aspects, the ChAdV68 vector comprises a ChAdV68 vector backbone comprising one or more deletions between base pair number 577 and 3403 or between base pair 456 and 3014, and optionally wherein the vector further comprises one or more deletions between base pair 27,125 and 31,825 or between base pair 27,816 and 31,333 of the sequence set forth in SEQ ID NO:1. In some aspects, the ChAdV68 vector comprises a ChAdV68 vector backbone comprising one or more deletions between base pair number 3957 and 10346, base pair number 21787 and 23370, and base pair number 33486 and 36193 of the sequence set forth in SEQ ID NO:1. In some aspects, the cassette is inserted in the ChAdV vector backbone at the E1 region, E3 region, and/or any deleted AdV region that allows incorporation of the cassette.

[0037] In some aspects, the at least one promoter nucleotide sequence is the native 26S promoter nucleotide sequence encoded by the backbone. In some aspects, the at least one promoter nucleotide sequence is an exogenous RNA promoter. In some aspects, the second promoter nucleotide sequence is a 26S promoter nucleotide sequence. In some aspects, the second promoter nucleotide sequence comprises multiple 26S promoter nucleotide sequences, wherein each 26S promoter nucleotide sequence provides for transcription of one or more of the separate open reading frames.

[0038] In some aspects, the one or more vectors are each at least 300nt in size. In some aspects, the one or more vectors are each at least 1kb in size. In some aspects, the one or more vectors are each 2kb in size. In some aspects, the one or more vectors are each less than 5kb in size.

[0039] In some aspects, at least one of the at least one antigen-encoding nucleic acid sequences encodes a polypeptide sequence or portion thereof that is presented by MHC class I on a cell surface, optionally a tumor cell surface or an infected cell surface.

[0040] In some aspects, each epitope-encoding nucleic acid sequence is linked directly to one another. In some aspects, at least one of the epitope-encoding nucleic acid sequences is linked to a distinct epitope-encoding nucleic acid sequence with a nucleic acid sequence encoding a linker. In some aspects, the linker links two MHC class I sequences or an MHC class

I sequence to an MHC class II sequence. In some aspects, the linker is selected from the group consisting of: (1) consecutive glycine residues, at least 2, 3, 4, 5, 6, 7, 8, 9, or 10 residues in length; (2) consecutive alanine residues, at least 2, 3, 4, 5, 6, 7, 8, 9, or 10 residues in length; (3) two arginine residues (RR); (4) alanine, alanine, tyrosine (AAY); (5) a consensus sequence at least 2, 3, 4, 5, 6, 7, 8, 9, or 10 amino acid residues in length that is processed efficiently by a mammalian proteasome; and (6) one or more native sequences flanking the antigen derived from the cognate protein of origin and that is at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, or 2-20 amino acid residues in length. In some aspects, the linker links two MHC class II sequences or an MHC class II sequence to an MHC class I sequence. In some aspects, the linker comprises the sequence GPGPG. In some aspects, at least one sequence of the epitope-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 sequences of epitope encoded therefrom. In some aspects, the separate or contiguous sequence comprises at least one of: a ubiquitin sequence, a ubiquitin sequence modified to increase proteasome targeting (e.g., the ubiquitin sequence contains a Gly to Ala substitution at position 76), an immunoglobulin signal sequence (e.g., IgK), a major histocompatibility class I sequence, lysosomal-associated membrane protein (LAMP)-1, human dendritic cell lysosomal-associated membrane protein, and a major histocompatibility class II sequence; optionally wherein the ubiquitin sequence modified to increase proteasome targeting is A76.

[0041] In some aspects, at least one of the epitope-encoding nucleic acid sequences encodes a polypeptide sequence or portion thereof that has increased binding affinity to its corresponding MHC allele relative to the translated, corresponding wild-type nucleic acid sequence. In some aspects, at least one of the epitope-encoding nucleic acid sequences encodes a polypeptide sequence or portion thereof that has increased binding stability to its corresponding MHC allele relative to the translated, corresponding wild-type nucleic acid sequence. In some aspects, at least one of the epitope-encoding nucleic acid sequences encodes a polypeptide sequence or portion thereof that has an increased likelihood of presentation on its corresponding MHC allele relative to the translated, corresponding wild-type nucleic acid sequence. In some aspects, the at least one alteration comprises a point mutation, a frameshift mutation, a non-frameshift mutation, a deletion mutation, an insertion mutation, a splice variant, a genomic rearrangement, or a proteasome-generated spliced antigen.

[0042] In some aspects, the tumor 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, or the infectious disease organism is selected from the group consisting of: Severe acute respiratory syndrome-related coronavirus (SARS), severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), Ebola, HIV, Hepatitis B virus (HBV), influenza, Hepatitis C virus (HCV), Human papillomavirus (HPV), Cytomegalovirus (CMV), Chikungunya virus, Respiratory syncytial virus (RSV), Dengue virus, an orthomyxoviridae family virus, and tuberculosis.

[0043] 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 epitope-encoding nucleic acid sequences. 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 epitope-encoding nucleic acid sequences. In some aspects, the at least one antigen-encoding nucleic acid sequence comprises at least 2-400 epitope-encoding nucleic acid sequences and wherein at least two of the epitope-encoding nucleic acid sequences encode polypeptide sequences or portions thereof that are presented by MHC class I on a cell surface, optionally a tumor cell surface or an infected cell surface. 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. 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 polypeptide sequences or portions thereof that are presented by MHC class I on a cell surface, optionally a tumor cell surface or an infected cell surface. In some aspects, at least two of the epitope-encoding nucleic acid sequences encode polypeptide sequences or portions thereof that are presented by MHC class I on a cell surface, optionally a tumor cell surface or an infected cell surface.

[0044] In some aspects, when administered to the subject and translated, at least one of the epitopes encoded by the epitope-encoding nucleic acid sequences are presented on antigen presenting cells resulting in an immune response targeting at least one of the antigens on the tumor cell surface or the infected cell surface. In some aspects, the at least one antigen-encoding nucleic acid sequences when administered to the subject and translated, at least one of the MHC class I or class II epitopes are presented on antigen presenting cells resulting in an immune response targeting at least one of the epitopes on a tumor cell surface or the infected cell surface,

and optionally wherein the expression of each of the at least one antigen-encoding nucleic acid sequences is driven by the at least one promoter nucleotide sequence.

[0045] In some aspects, each epitope-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.

[0046] 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 peptide sequence distinct from the corresponding peptide sequence encoded by a wild-type nucleic acid sequence. In some aspects, the at least one MHC class II epitope-encoding nucleic acid sequence is 12-20, 12, 13, 14, 15, 16, 17, 18, 19, 20, or 20-40 amino acids in length. In some aspects, the at least one MHC class II epitope-encoding nucleic acid sequence is present and comprises at least one universal MHC class II antigen-encoding nucleic acid sequence, optionally wherein the at least one universal sequence comprises at least one of Tetanus toxoid and PADRE.

[0047] 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.

[0048] In some aspects, the at least one poly(A) sequence comprises a poly(A) sequence native to the backbone. In some aspects, the at least one poly(A) sequence comprises a poly(A) sequence exogenous to the backbone. In some aspects, the at least one poly(A) sequence is operably linked to at least one of the at least one antigen-encoding 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, or at least 100 consecutive A nucleotides. In some aspects, the at least one poly(A) sequence is at least 80 consecutive A nucleotides.

[0049] In some aspects, the cassette further comprises at least one of: an intron sequence, a woodchuck hepatitis virus posttranscriptional regulatory element (WPRE) sequence, an internal ribosome entry sequence (IRES) sequence, a nucleotide sequence encoding a 2A self-cleaving peptide sequence, a nucleotide sequence encoding a Furin cleavage site, or a sequence in the 5' or 3' non-coding region known to enhance the nuclear export, stability, or translation efficiency of mRNA that is operably linked to at least one of the at least one antigen-encoding nucleic acid sequences. In some aspects, the cassette further comprises a reporter gene, including but not limited to, green fluorescent protein (GFP), a GFP variant, secreted alkaline phosphatase, luciferase, a luciferase variant, or a detectable peptide or epitope. In some aspects, the detectable

peptide or epitope is selected from the group consisting of an HA tag, a Flag tag, a His-tag, or a V5 tag.

[0050] In some aspects, the one or more vectors further comprises one or more nucleic acid sequences encoding at least one immune modulator. In some aspects, the immune modulator is an anti-CTLA4 antibody or an antigen-binding fragment thereof, an anti-PD-1 antibody or an antigen-binding fragment thereof, an anti-PD-L1 antibody or an antigen-binding fragment thereof, an anti-4-1BB antibody or an antigen-binding fragment thereof, or an anti-OX-40 antibody or an antigen-binding fragment thereof. In some aspects, the antibody or antigen-binding fragment thereof is a Fab fragment, a Fab' fragment, a single chain Fv (scFv), a single domain antibody (sdAb) either as single specific or multiple specificities linked together (e.g., camelid antibody domains), or full-length single-chain antibody (e.g., full-length IgG with heavy and light chains linked by a flexible linker). In some aspects, the heavy and light chain sequences of the antibody are a contiguous sequence separated by either a self-cleaving sequence such as 2A or IRES; or the heavy and light chain sequences of the antibody are linked by a flexible linker such as consecutive glycine residues. In some aspects, the immune modulator is a cytokine. In some aspects, the cytokine is at least one of IL-2, IL-7, IL-12, IL-15, or IL-21 or variants thereof of each.

[0051] In some aspects, at least one 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 antigens; (b) inputting the peptide sequence of each antigen into a presentation model to generate a set of numerical likelihoods that each of the antigens 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 antigens based on the set of numerical likelihoods to generate a set of selected antigens which are used to generate the epitope-encoding nucleic acid sequences.

[0052] In some aspects, each of the 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 antigens; (b) inputting the peptide sequence of each antigen into a presentation model to generate a set of numerical likelihoods that each of the antigens 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 antigens based on the set of numerical likelihoods to generate a set of selected antigens which are used to generate the at least 20 epitope-encoding nucleic acid sequences. In some aspects, a number of the set of selected antigens 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. In some aspects, selecting the set of selected antigens comprises selecting antigens that have an increased likelihood of being presented on the cell surface relative to unselected antigens based on the presentation model, optionally wherein the selected antigens have been validated as being presented by one or more specific HLA alleles. In some aspects, selecting the set of selected antigens comprises selecting antigens that have an increased likelihood of being capable of inducing a tumor-specific or infectious disease-specific immune response in the subject relative to unselected antigens based on the presentation model. In some aspects, selecting the set of selected antigens comprises selecting antigens that have an increased likelihood of being capable of being presented to naïve T cells by professional antigen presenting cells (APCs) relative to unselected antigens based on the presentation model, optionally wherein the APC is a dendritic cell (DC). In some aspects, selecting the set of selected antigens comprises selecting antigens that have a decreased likelihood of being subject to inhibition via central or peripheral tolerance relative to unselected antigens based on the presentation model. In some aspects, selecting the set of selected antigens comprises selecting antigens that have a decreased likelihood of being capable of inducing an autoimmune response to normal tissue in the subject relative to unselected antigens based on the presentation model. 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.

[0053] In some aspects, the cassette comprises junctional epitope sequences formed by adjacent sequences in the cassette. In some aspects, at least one or each junctional epitope sequence has an affinity of greater than 500 nM for MHC. In some aspects, each junctional epitope sequence is non-self.

[0054] In some aspects, each of the MHC class I epitopes is predicted or validated to be capable of presentation by at least one HLA allele present in at least 5% of a population. In some aspects, each of the MHC class I epitopes is predicted or validated to be capable of presentation

by at least one HLA allele, wherein each antigen/HLA pair has an antigen/HLA prevalence of at least 0.01% in a population. In some aspects, each of the MHC class I epitopes is predicted or validated to be capable of presentation by at least one HLA allele, wherein each antigen/HLA pair has an antigen/HLA prevalence of at least 0.1% in a population.

[0055] In some aspects, the cassette does not encode a non-therapeutic MHC class I or class II epitope nucleic acid sequence comprising a translated, wild-type nucleic acid sequence, wherein the non-therapeutic epitope is predicted to be displayed on an MHC allele of the subject. In some aspects, the non-therapeutic predicted MHC class I or class II epitope sequence is a junctional epitope sequence formed by adjacent sequences in the cassette.

[0056] In some aspects, the prediction is based on presentation likelihoods generated by inputting sequences of the non-therapeutic epitopes into a presentation model.

[0057] In some aspects, an order of the at least one antigen-encoding nucleic acid sequences in the cassette is determined by a series of steps comprising: (a) generating a set of candidate cassette sequences corresponding to different orders of the at least one antigen-encoding nucleic acid sequences; (b) determining, for each candidate cassette sequence, a presentation score based on presentation of non-therapeutic epitopes in the candidate cassette sequence; and (c) selecting a candidate cassette sequence associated with a presentation score below a predetermined threshold as the cassette sequence for an antigen vaccine.

[0058] Also provided for herein is a pharmaceutical composition comprising any of the compositions described herein and a pharmaceutically acceptable carrier. In some aspects, the composition further comprises an adjuvant. In some aspects, the composition further comprises an immune modulator. In some aspects, the immune modulator is an anti-CTLA4 antibody or an antigen-binding fragment thereof, an anti-PD-1 antibody or an antigen-binding fragment thereof, an anti-PD-L1 antibody or an antigen-binding fragment thereof, an anti-4-1BB antibody or an antigen-binding fragment thereof, or an anti-OX-40 antibody or an antigen-binding fragment thereof.

[0059] Also provided for herein is an isolated nucleotide sequence or set of isolated nucleotide sequences comprising the cassette of any of the compositions described herein and one or more elements obtained from the sequence of SEQ ID NO:3 or SEQ ID NO:5, optionally wherein the one or more elements are selected from the group consisting of the sequences necessary for nonstructural protein-mediated amplification, the 26S promoter nucleotide sequence, the poly(A) sequence, and the nsP1-4 genes of the sequence set forth in SEQ ID NO:3 or SEQ ID NO:5, and optionally wherein the nucleotide sequence is cDNA. In some aspects, the sequence or set of isolated nucleotide sequences comprises the cassette of any of the above composition claims inserted at position 7544 of the sequence set forth in SEQ ID NO:6 or SEQ

ID NO:7. In some aspects, the composition further comprises: a) a T7 or SP6 RNA polymerase promoter nucleotide sequence 5' of the one or more elements obtained from the sequence of SEQ ID NO:3 or SEQ ID NO:5; and b) optionally, one or more restriction sites 3' of the poly(A) sequence. In some aspects, the cassette of any of the above composition claims is inserted at position 7563 of SEQ ID NO:8 or SEQ ID NO:9.

[0060] Also provided for herein is a vector or set of vectors comprising any of the nucleotide sequence described herein.

[0061] Also provided for herein is an isolated cell comprising any of the nucleotide sequences or set of isolated nucleotide sequences described herein, optionally wherein the cell is a BHK-21, CHO, HEK293 or variants thereof, 911, HeLa, A549, LP-293, PER.C6, or AE1-2a cell.

[0062] Also provided for herein is a kit comprising any of the compositions described herein and instructions for use.

[0063] In some aspects, any of the above compositions further comprise a nanoparticulate delivery vehicle. The nanoparticulate delivery vehicle, in some aspects, may be a lipid nanoparticle (LNP). In some aspects, the LNP comprises ionizable amino lipids. In some aspects, the ionizable amino lipids comprise MC3-like (dilinoleylmethyl- 4-dimethylaminobutyrate) molecules. In some aspects, the nanoparticulate delivery vehicle encapsulates the antigen expression system.

[0064] In some aspects, any of the above compositions further comprise a plurality of LNPs, wherein the LNPs comprise: the antigen expression system; a cationic lipid; a non-cationic lipid; and a conjugated lipid that inhibits aggregation of the LNPs, wherein at least about 95% of the LNPs in the plurality of LNPs either: have a non-lamellar morphology; or are electron-dense.

[0065] In some aspects, the non-cationic lipid is a mixture of (1) a phospholipid and (2) cholesterol or a cholesterol derivative.

[0066] In some aspects, the conjugated lipid that inhibits aggregation of the LNPs is a polyethyleneglycol (PEG)-lipid conjugate. In some aspects, the PEG-lipid conjugate is selected from the group consisting of: a PEG-diacylglycerol (PEG-DAG) conjugate, a PEG dialkylxypropyl (PEG-DAA) conjugate, a PEG-phospholipid conjugate, a PEG-ceramide (PEG-Cer) conjugate, and a mixture thereof. In some aspects the PEG-DAA conjugate is a member selected from the group consisting of: a PEG-didecyloxypropyl (C₁₀) conjugate, a PEG-dilauryloxypropyl (C₁₂) conjugate, a PEG-dimyristyloxypropyl (C₁₄) conjugate, a PEG-dipalmitoyloxypropyl (C₁₆) conjugate, a PEG-distearoyloxypropyl (C₁₈) conjugate, and a mixture thereof.

[0067] In some aspects, the antigen expression system is fully encapsulated in the LNPs.

[0068] In some aspects, the non-lamellar morphology of the LNPs comprises an inverse hexagonal (H_{II}) or cubic phase structure.

[0069] In some aspects, the cationic lipid comprises from about 10 mol % to about 50 mol % of the total lipid present in the LNPs. In some aspects, the cationic lipid comprises from about 20 mol % to about 50 mol % of the total lipid present in the LNPs. In some aspects, the cationic lipid comprises from about 20 mol % to about 40 mol % of the total lipid present in the LNPs.

[0070] In some aspects, the non-cationic lipid comprises from about 10 mol % to about 60 mol % of the total lipid present in the LNPs. In some aspects, the non-cationic lipid comprises from about 20 mol % to about 55 mol % of the total lipid present in the LNPs. In some aspects, the non-cationic lipid comprises from about 25 mol % to about 50 mol % of the total lipid present in the LNPs.

[0071] In some aspects, the conjugated lipid comprises from about 0.5 mol % to about 20 mol % of the total lipid present in the LNPs. In some aspects, the conjugated lipid comprises from about 2 mol % to about 20 mol % of the total lipid present in the LNPs. In some aspects, the conjugated lipid comprises from about 1.5 mol % to about 18 mol % of the total lipid present in the LNPs.

[0072] In some aspects, greater than 95% of the LNPs have a non-lamellar morphology. In some aspects, greater than 95% of the LNPs are electron dense.

[0073] In some aspects, any of the above compositions further comprise a plurality of LNPs, wherein the LNPs comprise: a cationic lipid comprising from 50 mol % to 65 mol % of the total lipid present in the LNPs; a conjugated lipid that inhibits aggregation of LNPs comprising from 0.5 mol % to 2 mol % of the total lipid present in the LNPs; and a non-cationic lipid comprising either: a mixture of a phospholipid and cholesterol or a derivative thereof, wherein the phospholipid comprises from 4 mol % to 10 mol % of the total lipid present in the LNPs and the cholesterol or derivative thereof comprises from 30 mol % to 40 mol % of the total lipid present in the LNPs; a mixture of a phospholipid and cholesterol or a derivative thereof, wherein the phospholipid comprises from 3 mol % to 15 mol % of the total lipid present in the LNPs and the cholesterol or derivative thereof comprises from 30 mol % to 40 mol % of the total lipid present in the LNPs; or up to 49.5 mol % of the total lipid present in the LNPs and comprising a mixture of a phospholipid and cholesterol or a derivative thereof, wherein the cholesterol or derivative thereof comprises from 30 mol % to 40 mol % of the total lipid present in the LNPs.

[0074] In some aspects, any of the above compositions further comprise a plurality of LNPs, wherein the LNPs comprise: a cationic lipid comprising from 50 mol % to 85 mol % of the total lipid present in the LNPs; a conjugated lipid that inhibits aggregation of LNPs comprising from

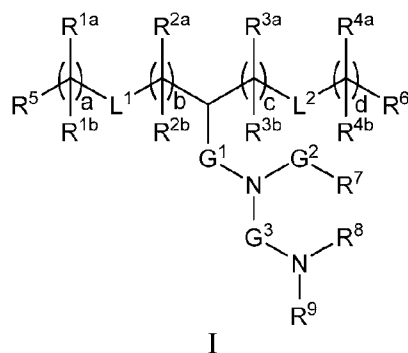
0.5 mol % to 2 mol % of the total lipid present in the LNPs; and a non-cationic lipid comprising from 13 mol % to 49.5 mol % of the total lipid present in the LNPs.

[0075] In some aspects, the phospholipid comprises dipalmitoylphosphatidylcholine (DPPC), distearoylphosphatidylcholine (DSPC), or a mixture thereof.

[0076] In some aspects, the conjugated lipid comprises a polyethyleneglycol (PEG)-lipid conjugate. In some aspects, the PEG-lipid conjugate comprises a PEG-diacylglycerol (PEG-DAG) conjugate, a PEG-dialkylxypropyl (PEG-DAA) conjugate, or a mixture thereof. In some aspects, the PEG-DAA conjugate comprises a PEG-dimyristyloxypropyl (PEG-DMA) conjugate, a PEG-distearoxypropyl (PEG-DSA) conjugate, or a mixture thereof. In some aspects, the PEG portion of the conjugate has an average molecular weight of about 2,000 daltons.

[0077] In some aspects, the conjugated lipid comprises from 1 mol % to 2 mol % of the total lipid present in the LNPs.

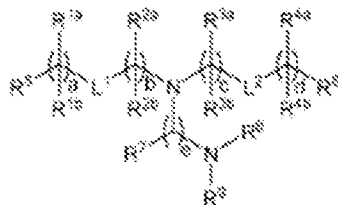
[0078] In some aspects, the LNP comprises a compound having a structure of Formula I:



or a pharmaceutically acceptable salt, tautomer, prodrug or stereoisomer thereof, wherein: L^1 and L^2 are each independently $-O(C=O)-$, $-(C=O)O-$, $-C(=O)-$, $-O-$, $-S(O)_x-$, $-S-S-$, $-C(=O)S-$, $-SC(=O)-$, $-R^aC(=O)-$, $-C(=O)R^a-$, $-R^aC(=O)R^a-$, $-OC(=O)R^a-$, $-R^aC(=O)O-$ or a direct bond; G^1 is $Ci-C_2$ alkylene, $-(C=O)-$, $-O(C=O)-$, $-SC(=O)-$, $-R^aC(=O)-$ or a direct bond; $-C(=O)-$, $-(C=O)O-$, $-C(=O)S-$, $-C(=O)R^a-$ or a direct bond; G is $Ci-C_6$ alkylene; R^a is H or C_1-C_{12} alkyl; R^{1a} and R^{1b} are, at each occurrence, independently either: (a) H or C_1-C_{12} alkyl; or (b) R^{1a} is H or C_1-C_{12} alkyl, and R^{1b} together with the carbon atom to which it is bound is taken together with an adjacent R^{1b} and the carbon atom to which it is bound to form a carbon-carbon double bond; R^{2a} and R^{2b} are, at each occurrence, independently either: (a) H or C_1-C_{12} alkyl; or (b) R^{2a} is H or C_1-C_{12} alkyl, and R^{2b} together with the carbon atom to which it is bound is taken together with an adjacent R^{2b} and the carbon atom to which it is bound to form a carbon-carbon double bond; R^{3a} and R^{3b} are, at each occurrence, independently either (a): H or C_1-C_{12} alkyl; or (b) R^{3a} is H or C_1-C_{12} alkyl, and R^{3b} together with the carbon atom to which it is bound is taken together with an adjacent R and the carbon atom to which it is bound to form a carbon-carbon double bond; R^{4a} and R^{4b} are, at each occurrence, independently either: (a) H or C_1-C_{12} alkyl; or (b) R^{4a} is H

or C1-C12 alkyl, and R^{4b} together with the carbon atom to which it is bound is taken together with an adjacent R^{4b} and the carbon atom to which it is bound to form a carbon-carbon double bond; R⁵ and R⁶ are each independently H or methyl; R⁷ is C4-C20 alkyl; R⁸ and R⁹ are each independently C1-C12 alkyl; or R⁸ and R⁹, together with the nitrogen atom to which they are attached, form a 5, 6 or 7-membered heterocyclic ring; a, b, c and d are each independently an integer from 1 to 24; and x is 0, 1 or 2.

[0079] In some aspects, the LNP comprises a compound having a structure of Formula II:



II

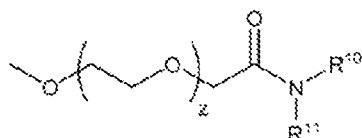
or a pharmaceutically acceptable salt, tautomer, prodrug or stereoisomer thereof, wherein: L¹ and L² are each independently -O(C=O)-, -(C=O)O- or a carbon-carbon double bond; R^{1a} and R^{1b} are, at each occurrence, independently either (a) H or C1-C12 alkyl, or (b) R^{1a} is H or C1-C12 alkyl, and R^{1b} together with the carbon atom to which it is bound is taken together with an adjacent R^{1b} and the carbon atom to which it is bound to form a carbon-carbon double bond; R^{2a} and R^{2b} are, at each occurrence, independently either (a) H or C1-C12 alkyl, or (b) R^{2a} is H or C1-C12 alkyl, and R^{2b} together with the carbon atom to which it is bound is taken together with an adjacent R^{2b} and the carbon atom to which it is bound to form a carbon-carbon double bond; R^{3a} and R^{3b} are, at each occurrence, independently either (a) H or C1-C12 alkyl, or (b) R^{3a} is H or C1-C12 alkyl, and R^{3b} together with the carbon atom to which it is bound is taken together with an adjacent R^{3b} and the carbon atom to which it is bound to form a carbon-carbon double bond; R^{4a} and R^{4b} are, at each occurrence, independently either (a) H or C1-C12 alkyl, or (b) R^{4a} is H or C1-C12 alkyl, and R^{4b} together with the carbon atom to which it is bound is taken together with an adjacent R^{4b} and the carbon atom to which it is bound to form a carbon-carbon double bond; R⁵ and R⁶ are each independently methyl or cycloalkyl; R⁷ is, at each occurrence, independently H or C1-C12 alkyl; R⁸ and R⁹ are each independently unsubstituted C1-C12 alkyl; or R⁸ and R⁹, together with the nitrogen atom to which they are attached, form a 5, 6 or 7-membered heterocyclic ring comprising one nitrogen atom; a and d are each independently an integer from 0 to 24; b and c are each independently an integer from 1 to 24; and e is 1 or 2, provided that: at least one of R^{1a}, R^{2a}, R^{3a} or R^{4a} is C1-C12 alkyl, or at least one of L¹ or L² is -O(C=O)- or -(C=O)O-; and R^{1a} and R^{1b} are not isopropyl when a is 6 or n-butyl when a is 8.

[0080] In some aspects, any of the above compositions further comprise one or more excipients comprising a neutral lipid, a steroid, and a polymer conjugated lipid. In some aspects, the neutral lipid comprises at least one of 1,2-Distearoyl-*sn*-glycero-3-phosphocholine (DSPC), 1,2-Dipalmitoyl-*sn*-glycero-3-phosphocholine (DPPC), 1,2-Dimyristoyl-*sn*-glycero-3-phosphocholine (DMPC), 1-Palmitoyl-2-oleoyl-*sn*-glycero-3-phosphocholine (POPC), 1,2-dioleoyl-*sn*-glycero-3-phosphocholine (DOPC), and 1,2-Dioleoyl-*sn*-glycero-3-phosphoethanolamine (DOPE). In some aspects, the neutral lipid is DSPC.

[0081] In some aspects, the molar ratio of the compound to the neutral lipid ranges from about 2:1 to about 8:1.

[0082] In some aspects, the steroid is cholesterol. In some aspects, the molar ratio of the compound to cholesterol ranges from about 2:1 to 1:1.

[0083] In some aspects, the polymer conjugated lipid is a pegylated lipid. In some aspects, the molar ratio of the compound to the pegylated lipid ranges from about 100:1 to about 25:1. In some aspects, the pegylated lipid is PEG-DAG, a PEG polyethylene (PEG-PE), a PEG-succinoyl-diacylglycerol (PEG-S-DAG), PEG-cer or a PEG dialkoxypyrpylcarbamate. In some aspects, the pegylated lipid has the following structure III:



III

or a pharmaceutically acceptable salt, tautomer or stereoisomer thereof, wherein: R^{10} and R^{11} are each independently a straight or branched, saturated or unsaturated alkyl chain containing from 10 to 30 carbon atoms, wherein the alkyl chain is optionally interrupted by one or more ester bonds; and z has a mean value ranging from 30 to 60. In some aspects, R^{10} and R^{11} are each independently straight, saturated alkyl chains having 12 to 16 carbon atoms. In some aspects, the average z is about 45.

start here

[0084] In some aspects, the LNP self-assembles into non-bilayer structures when mixed with polyanionic nucleic acid. In some aspects, the non-bilayer structures have a diameter between 60nm and 120nm. In some aspects, the non-bilayer structures have a diameter of about 70nm, about 80nm, about 90nm, or about 100nm. In some aspects, wherein the nanoparticulate delivery vehicle has a diameter of about 100nm.

[0085] Also provided for herein is a method for treating a subject with cancer, the method comprising administering to the subject any of the compositions or any of the pharmaceutical compositions described herein. In some aspects, the epitope-encoding nucleic acid sequence is

derived from the tumor of the subject with cancer or from a cell or sample of the infected subject. In some aspects, the epitope-encoding nucleic acid sequence are not derived from the tumor of the subject with cancer or from a cell or sample of the infected subject.

[0086] Also provided for herein is a method for stimulating an immune response in a subject, the method comprising administering to the subject any of the compositions or any of the pharmaceutical compositions described herein.

[0087] In some aspects, the subject expresses at least one HLA allele predicted or known to present the MHC class I epitope. In some aspects, HLA allele predicted or known to present the MHC class I epitope is A*03:01, A*11:01, A*02:01, C*01:02, and/or A*01:01. In some aspects, HLA allele predicted or known to present the MHC class I epitope is A*03:01. In some aspects, HLA allele predicted or known to present the MHC class I epitope is A*11:01. In some aspects, HLA allele predicted or known to present the MHC class I epitope is A*02:01. In some aspects, HLA allele predicted or known to present the MHC class I epitope is C*01:02. In some aspects, HLA allele predicted or known to present the MHC class I epitope is A*01:01. In some aspects, the composition is administered intramuscularly (IM), intradermally (ID), subcutaneously (SC), or intravenously (IV). In some aspects, the composition is administered intramuscularly. In some aspects, the method further comprising administration of one or more immune modulators, optionally wherein the immune modulator is administered before, concurrently with, or after administration of the composition or pharmaceutical composition. In some aspects, the one or more immune modulators are selected from the group consisting of: an anti-CTLA4 antibody or an antigen-binding fragment thereof, an anti-PD-1 antibody or an antigen-binding fragment thereof, an anti-PD-L1 antibody or an antigen-binding fragment thereof, an anti-4-1BB antibody or an antigen-binding fragment thereof, or an anti-OX-40 antibody or an antigen-binding fragment thereof. In some aspects, the immune modulator is administered intravenously (IV), intramuscularly (IM), intradermally (ID), or subcutaneously (SC). In some aspects, the subcutaneous administration is near the site of the composition or pharmaceutical composition administration or in close proximity to one or more vector or composition draining lymph nodes.

[0088] In some aspects, the method further comprises administering to the subject a second vaccine composition. In some aspects, the second vaccine composition is administered prior to the administration of any of the compositions or the pharmaceutical compositions described herein. In some aspects, the second vaccine composition is administered subsequent to the administration of any of the compositions or the pharmaceutical compositions described herein. In some aspects, the second vaccine composition is the same as any of the compositions or the pharmaceutical compositions described herein. In some aspects, the second vaccine composition is different any of the compositions or the pharmaceutical compositions described herein. In

some aspects, the second vaccine composition comprises a chimpanzee adenovirus vector encoding at least one antigen-encoding nucleic acid sequence. In some aspects, the at least one antigen-encoding nucleic acid sequence encoded by the chimpanzee adenovirus vector is the same as the at least one antigen-encoding nucleic acid sequence of any of the above composition claims.

[0089] Also provided for herein is a method of manufacturing the one or more vectors of any of the above composition claims, the method comprising: (a) obtaining a linearized DNA sequence comprising the backbone and the cassette; (b) *in vitro* transcribing the linearized DNA sequence by addition of the linearized DNA sequence to an *in vitro* transcription reaction containing all the necessary components to transcribe the linearized DNA sequence into RNA, optionally further comprising *in vitro* addition of the m7g cap to the resulting RNA; and (c) isolating the one or more vectors from the *in vitro* transcription reaction. In some aspects, the linearized DNA sequence is generated by linearizing a DNA plasmid sequence or by amplification using PCR. In some aspects, the DNA plasmid sequence is generated using one of bacterial recombination or full genome DNA synthesis or full genome DNA synthesis with amplification of synthesized DNA in bacterial cells. In some aspects, isolating the one or more vectors from the *in vitro* transcription reaction involves one or more of phenol chloroform extraction, silica column based purification, or similar RNA purification methods.

[0090] Also provided for herein is a method of manufacturing the composition of any of the above composition claims for delivery of the antigen expression system, the method comprising: (a) providing components for the nanoparticulate delivery vehicle; (b) providing the antigen expression system; and (c) providing conditions sufficient for the nanoparticulate delivery vehicle and the antigen expression system to produce the composition for delivery of the antigen expression system. In some aspects, the conditions are provided by microfluidic mixing.

[0091] Also provided for herein is a method for treating a subject with a disease, optionally wherein the disease is cancer or an infection, the method comprising administering to the subject an antigen-based vaccine to the subject, wherein the antigen-based vaccine comprises an antigen-encoding cassette, or a polypeptide sequence encoded by the cassette, wherein the antigen-encoding cassette comprises at least one antigen-encoding nucleic acid sequence described, from 5' to 3', by the formula:

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

wherein E represents a nucleotide sequence comprising a 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, and at least one of the distinct epitope-encoding nucleic acid sequences comprising the at least two iterations encodes a distinct KRAS-associated MHC class I neoepitope.

[0092] Also provided for herein is a method for treating a subject with a disease, optionally wherein the disease is cancer, the method comprising administering to the subject an antigen-based vaccine to the subject, wherein the antigen-based vaccine comprises an antigen expression system, comprising: the antigen expression system, wherein the antigen expression system comprises one or more vectors, the one or more vectors comprising: (a) a vector backbone, wherein the backbone comprises: (i) at least one promoter nucleotide sequence, and (ii) optionally, at least one polyadenylation (poly(A)) sequence; and (b) a cassette, wherein the cassette comprises: (i) at least one antigen-encoding nucleic acid sequence, comprising: (I) an epitope-encoding nucleic acid sequence encoding a KRAS-associated MHC class I neoepitope, and wherein each of the epitope-encoding nucleic acid sequences comprises: (A) optionally, a 5' linker sequence, and (B) optionally, a 3' linker sequence; (ii) optionally, a second promoter nucleotide sequence operably linked to the antigen-encoding nucleic acid sequence; and (iii) optionally, at least one MHC class II epitope-encoding nucleic acid sequence; (iv) optionally, at least one nucleic acid sequence encoding a GPGPS amino acid linker sequence (SEQ ID NO:56); and (v) optionally, at least one second poly(A) sequence, wherein the second poly(A) sequence is a native poly(A) sequence or an exogenous poly(A) sequence to the vector backbone, wherein if the second promoter nucleotide sequence is absent, the antigen-encoding nucleic acid sequence is operably linked to the at least one promoter nucleotide sequence, and wherein the at least one antigen-encoding nucleic acid sequence comprises at least two iterations of the epitope-encoding nucleic acid sequence encoding the KRAS-associated MHC class I neoepitope.

[0093] Also provided for herein is a method for treating a subject with a disease, optionally wherein the disease is cancer, the method comprising administering to the subject an antigen-based vaccine to the subject, wherein the antigen-based vaccine comprises an antigen expression system, comprising: the antigen expression system, wherein the antigen expression system comprises one or more vectors, the one or more vectors comprising: (a) a vector backbone, wherein the backbone comprises: (i) at least one promoter nucleotide sequence, and (ii) optionally, at least one polyadenylation (poly(A)) sequence; and (b) a cassette, wherein the cassette comprises: (i) at least one antigen-encoding nucleic acid sequence, comprising: (I) at least two distinct epitope-encoding nucleic acid sequences, 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, optionally wherein the at least one alteration is a KRAS mutation, 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, and wherein each of the epitope-encoding nucleic acid sequences comprises; (A) optionally, a 5' linker sequence, and (B) optionally, a 3' linker sequence; (ii) optionally, a second promoter nucleotide sequence operably linked to the antigen-encoding nucleic acid sequence; and (iii) optionally, at least one MHC class II epitope-encoding nucleic acid sequence; (iv) optionally, at least one nucleic acid sequence encoding a GPGPG amino acid linker sequence (SEQ ID NO:56); and (v) optionally, at least one second poly(A) sequence, wherein the second poly(A) sequence is a native poly(A) sequence or an exogenous poly(A) sequence to the vector backbone, wherein if the second promoter nucleotide sequence is absent, the antigen-encoding nucleic acid sequence is operably linked to the at least one promoter nucleotide sequence, and wherein the cassette does not encode an immunodominant MHC class I epitope that: (1) stimulates a 5-fold or greater immune response when administered in a vaccine composition to a subject relative to another MHC class I epitope encoded in the cassette and capable of stimulating an immune response in the subject, and/or (2) reduces an immune response to another MHC class I epitope encoded in the cassette when administered in a vaccine composition to a subject relative to an immune response when the other MHC class I epitope is administered in the absence of the immunodominant MHC class I epitope, optionally wherein the immune response is reduced to below a limit of detection and/or wherein the immune response is not a therapeutically effective response.

[0094] In some aspects, the antigen-encoding cassette encodes at least 4 iterations of each of the amino acid sequences VVVGACGVGK (SEQ ID NO: 75), VVVGADGVGK (SEQ ID NO: 78), VVGAVGVGK (SEQ ID NO: 79), and ILDTAGHEEY (SEQ ID NO: 82).

[0095] In some aspects, the epitope-encoding nucleic acid sequence is derived from a tumor of the subject with cancer or from a cell or sample of the infected subject. In some aspects, the epitope-encoding nucleic acid sequence are not derived from a tumor of the subject with cancer or from a cell or sample of the infected subject.

[0096] Also provided for herein is a method for stimulating an immune response in a subject, the method comprising the method comprising administering to the subject an antigen-based vaccine to the subject, wherein the antigen-based vaccine comprises an antigen-encoding cassette, or a polypeptide sequence encoded by the cassette, wherein the antigen-encoding

cassette comprises at least one antigen-encoding nucleic acid sequence described, from 5' to 3', by the formula:

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

wherein E represents a nucleotide sequence comprising a 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, and at least one of the distinct epitope-encoding nucleic acid sequences comprising the at least two iterations encodes a distinct KRAS-associated MHC class I neoepitope.

[0097] Also provided for herein is a method for stimulating an immune response in a subject, the method comprising the method comprising administering to the subject an antigen-based vaccine to the subject, wherein the antigen-based vaccine comprises: an antigen expression system, comprising: the antigen expression system, wherein the antigen expression system comprises one or more vectors, the one or more vectors comprising: (a) a vector backbone, wherein the backbone comprises: (i) at least one promoter nucleotide sequence, and (ii) optionally, at least one polyadenylation (poly(A)) sequence; and (b) a cassette, wherein the cassette comprises: (i) at least one antigen-encoding nucleic acid sequence, comprising: (I) an epitope-encoding nucleic acid sequence encoding a KRAS-associated MHC class I neoepitope, and wherein each of the epitope-encoding nucleic acid sequences comprises; (A) optionally, a 5' linker sequence, and (B) optionally, a 3' linker sequence; (ii) optionally, a second promoter nucleotide sequence operably linked to the antigen-encoding nucleic acid sequence; and (iii) optionally, at least one MHC class II epitope-encoding nucleic acid sequence; (iv) optionally, at least one nucleic acid sequence encoding a GP GPG amino acid linker sequence (SEQ ID NO:56); and (v) optionally, at least one second poly(A) sequence, wherein the second poly(A) sequence is a native poly(A) sequence or an exogenous poly(A) sequence to the vector backbone, wherein if the second promoter nucleotide sequence is absent, the antigen-encoding nucleic acid sequence is operably linked to the at least one promoter nucleotide sequence, and wherein the at least one antigen-encoding nucleic acid sequence comprises at least two iterations of at least one of the epitope-encoding nucleic acid sequences encoding the KRAS-associated MHC class I neoepitope.

[0098] Also provided for herein is a method for treating a subject with a disease, optionally wherein the disease is cancer or an infection, the method comprising administering to the subject

an antigen-based vaccine to the subject, wherein the antigen-based vaccine comprises an antigen expression system, comprising: the antigen expression system, wherein the antigen expression system comprises one or more vectors, the one or more vectors comprising: (a) a vector backbone, wherein the backbone comprises: (i) at least one promoter nucleotide sequence, and (ii) optionally, at least one polyadenylation (poly(A)) sequence; and (b) a cassette, wherein the cassette comprises: (i) at least one antigen-encoding nucleic acid sequence, comprising: (I) at least two distinct epitope-encoding nucleic acid sequences, 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, optionally wherein the at least one alteration is a KRAS mutation, 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, and wherein each of the epitope-encoding nucleic acid sequences comprises: (A) optionally, a 5' linker sequence, and (B) optionally, a 3' linker sequence; (ii) optionally, a second promoter nucleotide sequence operably linked to the antigen-encoding nucleic acid sequence; and (iii) optionally, at least one MHC class II epitope-encoding nucleic acid sequence; (iv) optionally, at least one nucleic acid sequence encoding a GPGPG amino acid linker sequence (SEQ ID NO:56); and (v) optionally, at least one second poly(A) sequence, wherein the second poly(A) sequence is a native poly(A) sequence or an exogenous poly(A) sequence to the vector backbone, wherein if the second promoter nucleotide sequence is absent, the antigen-encoding nucleic acid sequence is operably linked to the at least one promoter nucleotide sequence, and wherein the cassette does not encode an immunodominant MHC class I epitope that: (1) stimulates a 5-fold or greater immune response when administered in a vaccine composition to a subject relative to another MHC class I epitope encoded in the cassette and capable of stimulating an immune response in the subject, and/or (2) reduces an immune response to another MHC class I epitope encoded in the cassette when administered in a vaccine composition to a subject relative to an immune response when the other MHC class I epitope is administered in the absence of the immunodominant MHC class I epitope, optionally wherein the immune response is reduced to below a limit of detection and/or wherein the immune response is not a therapeutically effective response.

[0099] In some aspects, the antigen-encoding cassette encodes at least 4 iterations of each of the amino acid sequences VVVGACGVGK (SEQ ID NO: 75), VVVGADGVGK (SEQ ID NO: 78), VVGAVGVGK (SEQ ID NO: 79), and ILDTAGHEEY (SEQ ID NO: 82).

[00100] In some aspects, the subject expresses at least one HLA allele predicted or known to present the at least one epitope sequence, optionally wherein the at least one epitope sequence

predicted or known to be presented comprises (1) the KRAS-associated MHC class I neoepitope, and/or (2) the immunodominant MHC class I epitope and the other MHC class I epitope encoded in the cassette. In some aspects, the subject expresses at least one HLA allele predicted or known to present the at least one epitope sequence, and wherein the at least one epitope sequence comprises an epitope known or suspected to be presented by MHC class I on a surface of a cell, optionally wherein the at least one epitope sequence predicted or known to be presented comprises (1) the KRAS-associated MHC class I neoepitope, and/or (2) the immunodominant MHC class I epitope and the other MHC class I epitope encoded in the cassette. In some aspects, the surface of the cell is a tumor cell surface. In some aspects, the cell is a tumor cell 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, brain cancer, B-cell lymphoma, acute myelogenous leukemia, chronic myelogenous leukemia, chronic lymphocytic leukemia, T cell lymphocytic leukemia, non-small cell lung cancer, and small cell lung cancer. In some aspects, the surface of the cell is an infected cell surface. In some aspects, the cell is an infected cell selected from the group consisting of: a pathogen infected cell, a virally infected cell, a bacterially infected cell, a fungally infected cell, and a parasitically infected cell. In some aspects, the virally infected cell is selected from the group consisting of: an HIV infected cell, a Severe acute respiratory syndrome-related coronavirus (SARS) infected cell, a severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infected cell, a Ebola infected cell, a Hepatitis B virus (HBV) infected cell, an influenza infected cell, an orthomyxoviridae family virus infected cell, a Human papillomavirus (HPV) infected cell, a Cytomegalovirus (CMV) infected cell, a Chikungunya virus infected cell, a Respiratory syncytial virus (RSV) infected cell, a Dengue virus infected cell, and a Hepatitis C virus (HCV) infected cell.

[00101] Also provided for herein is a method for inducing an immune response in a subject, the method comprising administering to the subject an antigen-based vaccine to the subject, wherein the antigen-based vaccine comprises an antigen-encoding cassette, or a polypeptide sequence encoded by the cassette, wherein the antigen-encoding cassette comprises at least one antigen-encoding nucleic acid sequence described, from 5' to 3', by the formula:

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

wherein E represents a nucleotide sequence comprising a 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, and at least one of the distinct epitope-encoding nucleic acid sequences comprising the at least two iterations encodes a distinct KRAS-associated MHC class I neoepitope.

[00102] Also provided for herein is a method for inducing an immune response in a subject, the method comprising administering to the subject an antigen-based vaccine to the subject, wherein the antigen-based vaccine comprises: an antigen expression system, comprising: the antigen expression system, wherein the antigen expression system comprises one or more vectors, the one or more vectors comprising: (a) a vector backbone, wherein the backbone comprises: (i) at least one promoter nucleotide sequence, and (ii) optionally, at least one polyadenylation (poly(A)) sequence; and (b) a cassette, wherein the cassette comprises: (i) at least one antigen-encoding nucleic acid sequence, comprising: (I) an epitope-encoding nucleic acid sequence encoding a KRAS-associated MHC class I neoepitope, and wherein each of the epitope-encoding nucleic acid sequences comprises: (A) optionally, a 5' linker sequence, and (B) optionally, a 3' linker sequence; (ii) optionally, a second promoter nucleotide sequence operably linked to the antigen-encoding nucleic acid sequence; and (iii) optionally, at least one MHC class II epitope-encoding nucleic acid sequence; (iv) optionally, at least one nucleic acid sequence encoding a GPMPG amino acid linker sequence (SEQ ID NO:56); and (v) optionally, at least one second poly(A) sequence, wherein the second poly(A) sequence is a native poly(A) sequence or an exogenous poly(A) sequence to the vector backbone, wherein if the second promoter nucleotide sequence is absent, the antigen-encoding nucleic acid sequence is operably linked to the at least one promoter nucleotide sequence, wherein the at least one antigen-encoding nucleic acid sequence comprises at least two iterations of the epitope-encoding nucleic acid sequence encoding the KRAS-associated MHC class I neoepitope, and wherein the subject expresses at least one HLA allele predicted or known to present the at least one KRAS-associated MHC class I neoepitope.

[00103] Also provided for herein is a method for inducing an immune response in a subject, the method comprising administering to the subject an antigen-based vaccine to the subject, wherein the antigen-based vaccine comprises: an antigen expression system, comprising: the antigen expression system, wherein the antigen expression system comprises one or more vectors, the one or more vectors comprising: (a) a vector backbone, wherein the backbone comprises: (i) at least one promoter nucleotide sequence, and (ii) optionally, at least one polyadenylation (poly(A)) sequence; and (b) a cassette, wherein the cassette comprises: (i) at least one antigen-encoding nucleic acid sequence, comprising: (I) at least two 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, optionally wherein the at least one alteration is a KRAS mutation, 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, and optionally wherein the epitope-encoding nucleic acid sequence encodes a MHC class I epitope, and wherein each of the epitope-encoding nucleic acid sequences comprises; (A) optionally, a 5' linker sequence, and (B) optionally, a 3' linker sequence; (ii) optionally, a second promoter nucleotide sequence operably linked to the antigen-encoding nucleic acid sequence; and (iii) optionally, at least one MHC class II epitope-encoding nucleic acid sequence; (iv) optionally, at least one nucleic acid sequence encoding a GP GPG amino acid linker sequence (SEQ ID NO:56); and (v) optionally, at least one second poly(A) sequence, wherein the second poly(A) sequence is a native poly(A) sequence or an exogenous poly(A) sequence to the vector backbone, wherein if the second promoter nucleotide sequence is absent, the antigen-encoding nucleic acid sequence is operably linked to the at least one promoter nucleotide sequence, and wherein the cassette does not encode an immunodominant MHC class I epitope that: (1) stimulates a 5-fold or greater immune response when administered in a vaccine composition to a subject relative to another MHC class I epitope encoded in the cassette and capable of stimulating an immune response in the subject, and/or (2) reduces an immune response to another MHC class I epitope encoded in the cassette when administered in a vaccine composition to a subject relative to an immune response when the other MHC class I epitope is administered in the absence of the immunodominant MHC class I epitope, optionally wherein the immune response is reduced to below a limit of detection and/or wherein the immune response is not a therapeutically effective response, and wherein the subject expresses at least one HLA allele predicted or known to present both the immunodominant MHC class I epitope and the other MHC class I epitope encoded in the cassette.

[00104] In some aspects, the antigen-encoding cassette encodes at least 4 iterations of each of the amino acid sequences VVVGACGVGK (SEQ ID NO: 75), VVVGADGVGK (SEQ ID NO: 78), VVGAVGVGK (SEQ ID NO: 79), and ILDTAGHEEY (SEQ ID NO: 82). In some aspects, the KRAS-associated MHC class I neoepitope or the KRAS mutation comprises a KRAS G12C mutation, a KRAS G12V mutation, a KRAS G12D mutation, or a KRAS Q61H mutation. In some aspects, the KRAS-associated MHC class I neoepitope or the KRAS mutation comprises any one of the amino acid sequence shown in SEQ ID NOs: 75-82. In some aspects, the antigen-encoding cassette comprises each of the amino acid sequence shown in SEQ ID NOs: 75-82. In some aspects, the antigen-encoding cassette comprises two or more iterations of each of the amino acid sequence shown in SEQ ID NOs: 75-82. In some aspects, the antigen-encoding

cassette comprises 4 iterations of each of the amino acid sequence shown in SEQ ID NOs: 75-82. In some aspects, the KRAS-associated MHC class I neoepitope or the KRAS mutation comprises the amino acid sequence shown in SEQ ID NO: 57, SEQ ID NO: 58, SEQ ID NO: 59, or SEQ ID NO: 60. In some aspects, the epitope-encoding nucleic acid sequences comprises two or more distinct epitope-encoding nucleic acid sequences independently encoding a distinct KRAS-associated MHC class I neoepitope or a distinct KRAS mutation. In some aspects, each of the epitope-encoding nucleic acid sequences independently encodes a distinct KRAS-associated MHC class I neoepitope or a distinct KRAS mutation. In some aspects, the epitope-encoding nucleic acid sequences comprises two or more distinct epitope-encoding nucleic acid sequences independently encoding a KRAS G12C mutation, a KRAS G12V mutation, a KRAS G12D mutation, or a KRAS Q61H mutation. In some aspects, the epitope-encoding nucleic acid sequences independently encodes each of a KRAS G12C mutation, a KRAS G12V mutation, and a KRAS G12D mutation, and optionally a KRAS Q61H mutation. In some aspects, the antigen-encoding nucleic acid sequence encodes a peptide comprising the amino acid sequence shown in SEQ ID NO: 64 or SEQ ID NO: 65. In some aspects, the antigen-encoding nucleic acid sequence encodes a peptide comprising the amino acid sequence shown in SEQ ID NO: 65.

[00105] In some aspects, the cassette does not encode an immunodominant MHC class I epitope that: (1) stimulates a 5-fold or greater immune response when administered in a vaccine composition to a subject relative to another MHC class I epitope encoded in the cassette and capable of stimulating an immune response in the subject, and/or (2) reduces an immune response to another MHC class I epitope encoded in the cassette when administered in a vaccine composition to a subject relative to an immune response when the other MHC class I epitope is administered in the absence of the immunodominant MHC class I epitope, optionally wherein the immune response is reduced to below a limit of detection and/or wherein the immune response is not a therapeutically effective response. In some aspects, the cassette does not encode an immunodominant MHC class I epitope that stimulates a 5-fold or greater immune response when administered in a vaccine composition to a subject relative to a KRAS-associated neoepitope encoded in the cassette and capable of stimulating an immune response in the subject. In some aspects, the cassette does not encode an immunodominant MHC class I epitope that reduces an immune response to another MHC class I epitope encoded in the cassette when administered in a vaccine composition to a subject relative to an immune response when the other MHC class I epitope is administered in the absence of the immunodominant MHC class I epitope. In some aspects, the cassette does not encode an immunodominant MHC class I epitope that reduces an immune response to another MHC class I epitope encoded in the cassette to below a limit of detection when administered in a vaccine composition to a subject relative to an

immune response when the other MHC class I epitope is administered in the absence of the immunodominant MHC class I epitope. In some aspects, the cassette does not encode an immunodominant MHC class I epitope that reduces an immune response to another MHC class I epitope encoded in the cassette when administered in a vaccine composition to a subject relative to an immune response when the other MHC class I epitope is administered in the absence of the immunodominant MHC class I epitope, wherein the immune response to the other MHC class I epitope is not a therapeutically effective response.

[00106] In some aspects, the immunodominant epitope is a TP53-associated MHC class I neoepitope, optionally wherein the TP53-associated MHC class I neoepitope comprises a S127Y mutation.

[00107] In some aspects, the antigen expression system comprises any one of the antigen expression systems described herein. In some aspects, the antigen-based vaccine comprises any one of the pharmaceutical compositions described herein.

[00108] In some aspects, the antigen-based vaccine is administered as a priming dose. In some aspects, the antigen-based vaccine is administered as one or more boosting doses. In some aspects, the boosting dose is different than the priming dose. In some aspects, a) the priming dose comprises a chimpanzee adenovirus vector and the boosting dose comprises an alphavirus vector; or b) the priming dose comprises an alphavirus vector and the boosting dose comprises a chimpanzee adenovirus vector. In some aspects, the boosting dose is the same as the priming dose. In some aspects, the injection site of the one or more boosting doses is as close as possible to the injection site of the priming dose.

[00109] In some aspects, the method further comprises determining or having determined the HLA-haplotype of the subject.

[00110] In some aspects, the antigen-based vaccine is administered intramuscularly (IM), intradermally (ID), subcutaneously (SC), or intravenously (IV). In some aspects, the antigen-based vaccine is administered intramuscularly (IM). In some aspects, the IM administration is administered at separate injection sites. In some aspects, the separate injection sites are in opposing deltoid muscles. In some aspects, the separate injection sites are in gluteus or rectus femoris sites on each side.

[00111] Also disclosed herein is a pharmaceutical composition comprising any of the compositions disclosed herein (such as an alphavirus-based or ChAd-based vector disclosed herein) and a pharmaceutically acceptable carrier. In some aspects, the pharmaceutical composition further comprises an adjuvant. In some aspects, the pharmaceutical composition further comprises an immune modulator. In some aspects, the immune modulator is an anti-CTLA4 antibody or an antigen-binding fragment thereof, an anti-PD-1 antibody or an antigen-

binding fragment thereof, an anti-PD-L1 antibody or an antigen-binding fragment thereof, an anti-4-1BB antibody or an antigen-binding fragment thereof, or an anti-OX-40 antibody or an antigen-binding fragment thereof.

[00112] Also disclosed herein is a vector comprising an isolated nucleotide sequence disclosed herein.

[00113] Also disclosed herein is a kit comprising a vector or a composition disclosed herein and instructions for use.

[00114] Also disclosed herein is a method for treating a subject, the method comprising administering to the subject a vector disclosed herein or a pharmaceutical composition disclosed herein. Also disclosed herein is a method for inducing an immune response in a subject, the method comprising administering to the subject any of the compositions, vectors, or pharmaceutical compositions described herein. In some aspects, the subject expresses at least one HLA allele predicted or known to present the MHC class I epitope. In some aspects, HLA allele predicted or known to present the MHC class I epitope is A*03:01, A*11:01, A*02:01, C*01:02, and/or A*01:01. In some aspects, HLA allele predicted or known to present the MHC class I epitope is A*03:01. In some aspects, HLA allele predicted or known to present the MHC class I epitope is A*11:01. In some aspects, HLA allele predicted or known to present the MHC class I epitope is A*02:01. In some aspects, HLA allele predicted or known to present the MHC class I epitope is C*01:02. In some aspects, HLA allele predicted or known to present the MHC class I epitope is A*01:01. In some aspects, the vector or composition is administered intramuscularly (IM), intradermally (ID), or subcutaneously (SC), or intravenously (IV).

[00115] Also disclosed herein is a method of manufacturing the one or more vectors of any of the above compositions, the method comprising: obtaining a linearized DNA sequence comprising the backbone and the antigen cassette; *in vitro* transcribing the linearized DNA sequence by addition of the linearized DNA sequence to a *in vitro* transcription reaction containing all the necessary components to transcribe the linearized DNA sequence into RNA, optionally further comprising *in vitro* addition of the m7g cap to the resulting RNA; and isolating the one or more vectors from the *in vitro* transcription reaction. In some aspects, the linearized DNA sequence is generated by linearizing a DNA plasmid sequence or by amplification using PCR. In some aspects, the DNA plasmid sequence is generated using one of bacterial recombination or full genome DNA synthesis or full genome DNA synthesis with amplification of synthesized DNA in bacterial cells. In some aspects, the isolating the one or more vectors from the *in vitro* transcription reaction involves one or more of phenol chloroform extraction, silica column based purification, or similar RNA purification methods.

[00116] Also disclosed herein is a method of manufacturing any of the compositions disclosed herein, the method comprising: providing components for the nanoparticulate delivery vehicle; providing the antigen expression system; and providing conditions sufficient for the nanoparticulate delivery vehicle and the antigen expression system to produce the composition for delivery of the antigen expression system. In some aspects, the conditions are provided by microfluidic mixing.

[00117] Also disclosed herein is a method of manufacturing a adenovirus vector disclosed herein, the method comprising: obtaining a plasmid sequence comprising the at least one promoter sequence and the antigen cassette; transfecting the plasmid sequence into one or more host cells; and isolating the adenovirus vector from the one or more host cells.

[00118] In some aspects, isolating comprises: lysing the host cell to obtain a cell lysate comprising the adenovirus vector; and purifying the adenovirus vector from the cell lysate.

[00119] In some aspects, the plasmid sequence is generated using one of bacterial recombination or full genome DNA synthesis or full genome DNA synthesis with amplification of synthesized DNA in bacterial cells. In some aspects, the one or more host cells are at least one of CHO, HEK293 or variants thereof, 911, HeLa, A549, LP-293, PER.C6, and AE1-2a cells. In some aspects, purifying the adenovirus vector from the cell lysate involves one or more of chromatographic separation, centrifugation, virus precipitation, and filtration.

BRIEF DESCRIPTION OF THE SEVERAL VIEWS OF THE DRAWINGS

[00120] 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:

[00121] **Figure (FIG.) 1A** presents an illustration of cassettes featuring either a single copy of KRAS neoepitopes G12C, G12V, G12D, and Q61H (“KRAS 1X (20x1)”; cassette = SEQ ID NO: 63), 2 repeats of the KRAS G12C, G12V, G12D, and Q61H neoepitopes and 2 repeats of additional KRAS neoepitopes (“KRAS 2X (8x2)”; cassette = SEQ ID NO: 64), or 4 repeats of the KRAS neoepitopes (“KRAS 4X (4x4)”; cassette = SEQ ID NO: 65). Numerical identifiers are in reference to the epitope “slot” relative to each cassette respectively and not across cassette designs (*e.g.*, the slot “3” epitope in the 20x1 cassette is not the same as the epitope in slot 3 of the 8x2 cassette).

[00122] **FIG. 1B** demonstrates repeating epitopes increases vaccine induced antigen-specific T-cell response. Shown are ELISpot results for the repeated neoepitope KRAS G12C. Mice engineered to express human HLA-A11:01 were immunized with 8×10^{10} VP using the ChAdV68 delivery vectors indicated and splenocytes isolated 14 days post-immunization. The number of

antigen-specific T-cells were measured by IFN γ ELISpot following overnight stimulation with VVVGACGVGK (SEQ ID NO: 75). Data presented as spot forming colonies (SFC) per 1×10^6 splenocytes for each animal. Bar represents the median.

[00123] FIG. 1C demonstrates repeating epitopes increases vaccine induced antigen-specific T-cell response. Shown are ELISpot results for the repeated neopeptide KRAS G12V. Mice engineered to express human HLA-A11:01 were immunized with 8×10^{10} VP using the ChAdV68 delivery vectors indicated and splenocytes isolated 14 days post-immunization. The number of antigen-specific T-cells were measured by IFN γ ELISpot following overnight stimulation with VVVGAVGVGK (SEQ ID NO: 81). Data presented as spot forming colonies (SFC) per 1×10^6 splenocytes for each animal. Bar represents the median. Dashed line represent samples that were too numerous to count (TNTC).

[00124] FIG. 1D demonstrates repeating epitopes increases vaccine induced antigen-specific T-cell response. Shown are ELISpot results for the repeated neopeptide KRAS G12D. Mice engineered to express human HLA-A11:01 were immunized with 8×10^{10} VP using the ChAdV68 delivery vectors indicated and splenocytes isolated 14 days post-immunization. The number of antigen-specific T-cells were measured by IFN γ ELISpot following overnight stimulation with VVVGADGVGK (SEQ ID NO: 78). Data presented as spot forming colonies (SFC) per 1×10^6 splenocytes for each animal. Bar represents the median.

[00125] FIG. 2A presents an illustration of ChAdV68 delivery vectors designed to assess immunodominance of a TP53 epitope, specifically vectors containing only KRAS neopeptides G12C, G12V, G12D, and Q61H ("KRAS 4x1"; cassette = SEQ ID NO: 66), KRAS neopeptides in combination with a TP53 R213L neopeptide ("KRAS 4x1 + R213L"; cassette = SEQ ID NO: 67), and KRAS neopeptides in combination with a TP53 S127Y neopeptide ("KRAS 4x1 + S127Y"; cassette = SEQ ID NO: 68).

[00126] FIG. 2B demonstrates removal of an immunodominant epitope increases vaccine induced antigen-specific T-cell response to KRAS neopeptides. Shown are ELISpot results for the neopeptide KRAS G12C. Mice engineered to express human HLA-A11:01 were immunized with 5×10^{10} VP using the ChAdV68 delivery vectors indicated and splenocytes isolated 14 days post-immunization. The number of antigen-specific T-cells were measured by IFN γ ELISpot following overnight stimulation with VVVGACGVGK (SEQ ID NO: 75). Data presented as spot forming colonies (SFC) per 1×10^6 splenocytes for each animal. Bar represents the median.

[00127] FIG. 2C demonstrates removal of an immunodominant epitope increases vaccine induced antigen-specific T-cell response to KRAS neopeptides. Shown are ELISpot results for the neopeptide KRAS G12D. Mice engineered to express human HLA-A11:01 were immunized with 5×10^{10} VP using the ChAdV68 delivery vectors indicated and splenocytes isolated 14 days

post-immunization. The number of antigen-specific T-cells were measured by IFN γ ELISpot following overnight stimulation with VVVGADGVGK (SEQ ID NO: 78). Data presented as spot forming colonies (SFC) per 1×10^6 splenocytes for each animal. Bar represents the median.

[00128] FIG. 2D demonstrates removal of an immunodominant epitope increases vaccine induced antigen-specific T-cell response to KRAS neoepitopes. Shown are ELISpot results for the neoepitope KRAS G12V. Mice engineered to express human HLA-A11:01 were immunized with 5×10^{10} VP using the ChAdV68 delivery vectors indicated and splenocytes isolated 14 days post-immunization. The number of antigen-specific T-cells were measured by IFN γ ELISpot following overnight stimulation with VVVGAVGVGK (SEQ ID NO: 81). Data presented as spot forming colonies (SFC) per 1×10^6 splenocytes for each animal. Bar represents the median.

[00129] FIG. 2E demonstrates the immune response of an immunodominant epitope and related control epitope. Shown are ELISpot results for the TP53 neoepitope pools for R213L and S127Y neoepitopes. Mice engineered to express human HLA-A11:01 were immunized with 5×10^{10} VP using the ChAdV68 delivery vectors indicated and splenocytes isolated 14 days post-immunization. The number of antigen-specific T-cells were measured by IFN γ ELISpot following overnight stimulation. Data presented as spot forming colonies (SFC) per 1×10^6 splenocytes for each animal. Bar represents the median. Dashed line represent samples that were too numerous to count (TNTC).

[00130] FIG. 3 demonstrates repeating epitopes increases vaccine induced antigen-specific T-cell response. Shown are ELISpot results for the repeated neoepitope KRAS G12V (left panel) or KRAS G12D (right panel). Mice engineered to express human HLA-A11:01 were immunized with 5×10^{10} VP using the ChAdV68 delivery vectors indicated and splenocytes isolated 14 days post-immunization. The number of antigen-specific T-cells were measured by IFN γ ELISpot following overnight stimulation with VVVGAVGVGK (SEQ ID NO: 81) or VVVGADGVGK (SEQ ID NO: 78), respectively. Data presented as spot forming colonies (SFC) per 1×10^6 splenocytes for each animal. Bar represents the median.

[00131] FIG. 4 demonstrates repeating epitopes increases vaccine induced antigen-specific T-cell response. Shown are ELISpot results for the repeated neoepitope KRAS G12V (left panel) or KRAS G12D (right panel). Mice engineered to express human HLA-A11:01 were immunized with 7×10^{10} VP using the ChAdV68 delivery vectors indicated and splenocytes isolated 14 days post-immunization. The number of antigen-specific T-cells were measured by IFN γ ELISpot following overnight stimulation with VVVGAVGVGK (SEQ ID NO: 81) or VVVGADGVGK (SEQ ID NO: 78), respectively. Data presented as spot forming colonies (SFC) per 1×10^6 splenocytes for each animal. Bar represents the median.

[00132] FIG. 5 demonstrates repeating epitopes increases vaccine induced antigen-specific T-cell response for KRAS Q61H. Shown are ELISpot results for the repeated neoepitope KRAS Q61H for the indicated cassette formats. Mice engineered to express human HLA-A01:01 were immunized with 5×10^{10} VP using the ChAdV68 delivery vectors indicated and splenocytes isolated 12 days post-immunization. The number of antigen-specific T-cells were measured by IFN γ ELISpot following overnight stimulation with ILDTAGHEEY (SEQ ID NO: 82). Data presented as spot forming colonies (SFC) per 1×10^6 splenocytes for each animal. Bar represents the median. Dashed line represent samples that were too numerous to count (TNTC).

[00133] FIG. 6 demonstrates repeating epitopes increases vaccine induced antigen-specific T-cell response for both ChAdV68 and SAM vector formats. Shown are ELISpot results for the repeated neoepitope KRAS G12V (left panel) or KRAS G12D (right panel). Mice engineered to express human HLA-A11:01 were immunized with 5×10^{10} VP using the ChAdV68 delivery vectors indicated or 10 μ g the SAM vectors indicated and splenocytes isolated 14 days post-immunization. The number of antigen-specific T-cells were measured by IFN γ ELISpot following overnight stimulation with respective peptide pools that contained all possible 38 minimal epitopes that span the 25mer. Data presented as spot forming colonies (SFC) per 1×10^6 splenocytes for each animal. Bar represents the median. Columns from left to right are ChAdV68 20x1, SAM 20x1, ChAdV68 4x4, and SAM 4x4.

[00134] FIG. 7 demonstrates repeating epitopes increases vaccine induced antigen-specific T-cell response for both ChAdV68 and SAM vector formats. Shown are ELISpot results for the repeated neoepitope KRAS G12V (left panel) or KRAS G12D (right panel). Mice engineered to express human HLA-A11:01 were immunized with 5×10^{10} VP using the ChAdV68 delivery vectors indicated or 10 μ g the SAM vectors indicated and splenocytes isolated 14 days post-immunization. The number of antigen-specific T-cells were measured by IFN γ ELISpot following overnight stimulation with VVGAVGVGK (SEQ ID NO: 79) or VVVGADGVGK (SEQ ID NO: 78). Data presented as spot forming colonies (SFC) per 1×10^6 splenocytes for each animal. Bar represents the median. Columns from left to right are ChAdV68 20x1, SAM 20x1, ChAdV68 4x4, and SAM 4x4.

DETAILED DESCRIPTION

I. Definitions

[00135] In general, terms used in the claims and the specification are intended to be construed as having the plain meaning understood by a person of ordinary skill in the art. Certain terms are defined below to provide additional clarity. In case of conflict between the plain meaning and the provided definitions, the provided definitions are to be used.

[00136] As used herein the term “antigen” is a substance that stimulates an immune response. An antigen can be a neoantigen. An antigen can be a “shared antigen” that is an antigen found among a specific population, *e.g.*, a specific population of cancer patients.

[00137] As used herein the term “neoantigen” is an antigen that has at least one alteration that makes it distinct from the corresponding wild-type antigen, *e.g.*, via mutation in a tumor cell or post-translational modification specific to a tumor cell. A neoantigen can include a polypeptide sequence or a nucleotide sequence. A mutation can include a frameshift or non-frameshift indel, missense or nonsense substitution, splice site alteration, genomic rearrangement or gene fusion, or any genomic or expression alteration giving rise to a neoORF. A mutations can also include a splice variant. Post-translational modifications specific to a tumor cell can include aberrant phosphorylation. Post-translational modifications specific to a tumor cell can also include a proteasome-generated spliced antigen. *See* Liepe et al., A large fraction of HLA class I ligands are proteasome-generated spliced peptides; *Science*. 2016 Oct 21;354(6310):354-358. The subject can be identified for administration through the use of various diagnostic methods, *e.g.*, patient selection methods described further below.

[00138] As used herein the term “tumor antigen” is an antigen present in a subject’s tumor cell or tissue but not in the subject’s corresponding normal cell or tissue, or derived from a 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.

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

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

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

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

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

[00144] As used herein the term “NEO-ORF” is a tumor-specific ORF arising from a mutation or other aberration such as splicing.

[00145] As used herein the term “missense mutation” is a mutation causing a substitution from one amino acid to another.

[00146] As used herein the term “nonsense mutation” is a mutation causing a substitution from an amino acid to a stop codon or causing removal of a canonical start codon.

[00147] As used herein the term “frameshift mutation” is a mutation causing a change in the frame of the protein.

[00148] As used herein the term “indel” is an insertion or deletion of one or more nucleic acids.

[00149] As used herein, the term percent "identity," in the context of two or more nucleic acid or polypeptide sequences, refer to two or more sequences or subsequences that have a specified percentage of nucleotides or amino acid residues that are the same, when compared and aligned for maximum correspondence, as measured using one of the sequence comparison algorithms described below (e.g., BLASTP and BLASTN or other algorithms available to persons of skill) or by visual inspection. Depending on the application, the percent "identity" can exist over a region of the sequence being compared, e.g., over a functional domain, or, alternatively, exist over the full length of the two sequences to be compared.

[00150] For sequence comparison, typically one sequence acts as a reference sequence to which test sequences are compared. When using a sequence comparison algorithm, test and reference sequences are input into a computer, subsequence coordinates are designated, if necessary, and sequence algorithm program parameters are designated. The sequence comparison algorithm then calculates the percent sequence identity for the test sequence(s) relative to the reference sequence, based on the designated program parameters. Alternatively, sequence similarity or dissimilarity can be established by the combined presence or absence of particular nucleotides, or, for translated sequences, amino acids at selected sequence positions (e.g., sequence motifs).

[00151] Optimal alignment of sequences for comparison can be conducted, e.g., by the local homology algorithm of Smith & Waterman, *Adv. Appl. Math.* 2:482 (1981), by the homology alignment algorithm of Needleman & Wunsch, *J. Mol. Biol.* 48:443 (1970), by the search for similarity method of Pearson & Lipman, *Proc. Nat'l. Acad. Sci. USA* 85:2444 (1988), by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, Wis.), or by visual inspection (see generally Ausubel et al., *infra*).

[00152] One example of an algorithm that is suitable for determining percent sequence identity and sequence similarity is the BLAST algorithm, which is described in Altschul et al., *J. Mol. Biol.* 215:403-410 (1990). Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information.

[00153] As used herein the term “non-stop or read-through” is a mutation causing the removal of the natural stop codon.

[00154] As used herein the term “epitope” is the specific portion of an antigen typically bound by an antibody or T cell receptor.

[00155] As used herein the term “immunogenic” is the ability to stimulate an immune response, e.g., via T cells, B cells, or both.

[00156] As used herein the term “HLA binding affinity” “MHC binding affinity” means affinity of binding between a specific antigen and a specific MHC allele.

[00157] As used herein the term “bait” is a nucleic acid probe used to enrich a specific sequence of DNA or RNA from a sample.

[00158] As used herein the term “variant” is a difference between a subject’s nucleic acids and the reference human genome used as a control.

[00159] As used herein the term “variant call” is an algorithmic determination of the presence of a variant, typically from sequencing.

[00160] As used herein the term “polymorphism” is a germline variant, i.e., a variant found in all DNA-bearing cells of an individual.

[00161] As used herein the term “somatic variant” is a variant arising in non-germline cells of an individual.

[00162] As used herein the term “allele” is a version of a gene or a version of a genetic sequence or a version of a protein.

[00163] As used herein the term “HLA type” is the complement of HLA gene alleles.

[00164] As used herein the term “nonsense-mediated decay” or “NMD” is a degradation of an mRNA by a cell due to a premature stop codon.

[00165] As used herein the term “truncal mutation” is a mutation originating early in the development of a tumor and present in a substantial portion of the tumor’s cells.

[00166] As used herein the term “subclonal mutation” is a mutation originating later in the development of a tumor and present in only a subset of the tumor’s cells.

[00167] As used herein the term “exome” is a subset of the genome that codes for proteins. An exome can be the collective exons of a genome.

[00168] As used herein the term “logistic regression” is a regression model for binary data from statistics where the logit of the probability that the dependent variable is equal to one is modeled as a linear function of the dependent variables.

[00169] As used herein the term “neural network” is a machine learning model for classification or regression consisting of multiple layers of linear transformations followed by element-wise nonlinearities typically trained via stochastic gradient descent and back-propagation.

[00170] As used herein the term “proteome” is the set of all proteins expressed and/or translated by a cell, group of cells, or individual.

[00171] As used herein the term “peptidome” is the set of all peptides presented by MHC-I or MHC-II on the cell surface. The peptidome may refer to a property of a cell or a collection of cells (e.g., the tumor peptidome, meaning the union of the peptidomes of all cells that comprise the tumor, or the infectious disease peptidome, meaning the union of the peptidomes of all cells that are infected by the infectious disease).

[00172] As used herein the term “ELISPOT” means Enzyme-linked immunosorbent spot assay – which is a common method for monitoring immune responses in humans and animals.

[00173] As used herein the term “dextramers” is a dextran-based peptide-MHC multimers used for antigen-specific T-cell staining in flow cytometry.

[00174] As used herein the term “tolerance or immune tolerance” is a state of immune non-responsiveness to one or more antigens, e.g. self-antigens.

[00175] As used herein the term “central tolerance” is a tolerance affected in the thymus, either by deleting self-reactive T-cell clones or by promoting self-reactive T-cell clones to differentiate into immunosuppressive regulatory T-cells (Tregs).

[00176] As used herein the term “peripheral tolerance” is a tolerance affected in the periphery by downregulating or anergizing self-reactive T-cells that survive central tolerance or promoting these T cells to differentiate into Tregs.

[00177] The term “sample” can include a single cell or multiple cells or fragments of cells or an aliquot of body fluid, taken from a subject, by means including venipuncture, excretion, ejaculation, massage, biopsy, needle aspirate, lavage sample, scraping, surgical incision, or intervention or other means known in the art.

[00178] The term “subject” encompasses a cell, tissue, or organism, human or non-human, whether in vivo, ex vivo, or in vitro, male or female. The term subject is inclusive of mammals including humans.

[00179] The term “mammal” encompasses both humans and non-humans and includes but is not limited to humans, non-human primates, canines, felines, murines, bovines, equines, and porcines.

[00180] The term “clinical factor” refers to a measure of a condition of a subject, e.g., disease activity or severity. “Clinical factor” encompasses all markers of a subject’s health status, including non-sample markers, and/or other characteristics of a subject, such as, without limitation, age and gender. A clinical factor can be a score, a value, or a set of values that can be obtained from evaluation of a sample (or population of samples) from a subject or a subject under a determined condition. A clinical factor can also be predicted by markers and/or other

parameters such as gene expression surrogates. Clinical factors can include tumor type, tumor sub-type, infection type, infection sub-type, and smoking history.

[00181] The term “antigen-encoding nucleic acid sequences derived from a tumor” refers to nucleic acid sequences obtained from the tumor, e.g. via RT-PCR; or sequence data obtained by sequencing the tumor and then synthesizing the nucleic acid sequences using the sequencing data, e.g., via various synthetic or PCR-based methods known in the art. Derived sequences can include nucleic acid sequence variants, such as sequence-optimized nucleic acid sequence variants (e.g., codon-optimized and/or otherwise optimized for expression), that encode the same polypeptide sequence as the corresponding native nucleic acid sequence obtained from a tumor.

[00182] The term “antigen-encoding nucleic acid sequences derived from an infection” refers to nucleic acid sequences obtained from infected cells or an infectious disease organism, e.g. via RT-PCR; or sequence data obtained by sequencing the infected cell or infectious disease organism and then synthesizing the nucleic acid sequences using the sequencing data, e.g., via various synthetic or PCR-based methods known in the art. Derived sequences can include nucleic acid sequence variants, such as sequence-optimized nucleic acid sequence variants (e.g., codon-optimized and/or otherwise optimized for expression), that encode the same polypeptide sequence as the corresponding native infectious disease organism nucleic acid sequence. Derived sequences can include nucleic acid sequence variants that encode a modified infectious disease organism polypeptide sequence having one or more (e.g., 1, 2, 3, 4, or 5) mutations relative to a native infectious disease organism polypeptide sequence. For example, a modified polypeptide sequence can have one or more missense mutations relative to the native polypeptide sequence of an infectious disease organism protein.

[00183] The term “alphavirus” refers to members of the family *Togaviridae*, and are positive-sense single-stranded RNA viruses. Alphaviruses are typically classified as either Old World, such as Sindbis, Ross River, Mayaro, Chikungunya, and Semliki Forest viruses, or New World, such as eastern equine encephalitis, Aura, Fort Morgan, or Venezuelan equine encephalitis and its derivative strain TC-83. Alphaviruses are typically self-replicating RNA viruses.

[00184] The term “alphavirus backbone” refers to minimal sequence(s) of an alphavirus that allow for self-replication of the viral genome. Minimal sequences can include conserved sequences for nonstructural protein-mediated amplification, a nonstructural protein 1 (nsP1) gene, a nsP2 gene, a nsP3 gene, a nsP4 gene, and a polyA sequence, as well as sequences for expression of subgenomic viral RNA including a subgenomic (e.g., a 26S) promoter element.

[00185] The term “sequences for nonstructural protein-mediated amplification” includes alphavirus conserved sequence elements (CSE) well known to those in the art. CSEs include, but

are not limited to, an alphavirus 5' UTR, a 51-nt CSE, a 24-nt CSE, a subgenomic promoter sequence (*e.g.*, a 26S subgenomic promoter sequence), a 19-nt CSE, and an alphavirus 3' UTR.

[00186] The term “RNA polymerase” includes polymerases that catalyze the production of RNA polynucleotides from a DNA template. RNA polymerases include, but are not limited to, bacteriophage derived polymerases including T3, T7, and SP6.

[00187] The term “lipid” includes hydrophobic and/or amphiphilic molecules. Lipids can be cationic, anionic, or neutral. Lipids can be synthetic or naturally derived, and in some instances biodegradable. Lipids can include cholesterol, phospholipids, lipid conjugates including, but not limited to, polyethyleneglycol (PEG) conjugates (PEGylated lipids), waxes, oils, glycerides, fats, and fat-soluble vitamins. Lipids can also include dilinoleylmethyl- 4-dimethylaminobutyrate (MC3) and MC3-like molecules.

[00188] The term “lipid nanoparticle” or “LNP” includes vesicle like structures formed using a lipid containing membrane surrounding an aqueous interior, also referred to as liposomes. Lipid nanoparticles includes lipid-based compositions with a solid lipid core stabilized by a surfactant. The core lipids can be fatty acids, acylglycerols, waxes, and mixtures of these surfactants. Biological membrane lipids such as phospholipids, sphingomyelins, bile salts (sodium taurocholate), and sterols (cholesterol) can be utilized as stabilizers. Lipid nanoparticles can be formed using defined ratios of different lipid molecules, including, but not limited to, defined ratios of one or more cationic, anionic, or neutral lipids. Lipid nanoparticles can encapsulate molecules within an outer-membrane shell and subsequently can be contacted with target cells to deliver the encapsulated molecules to the host cell cytosol. Lipid nanoparticles can be modified or functionalized with non-lipid molecules, including on their surface. Lipid nanoparticles can be single-layered (unilamellar) or multi-layered (multilamellar). Lipid nanoparticles can be complexed with nucleic acid. Unilamellar lipid nanoparticles can be complexed with nucleic acid, wherein the nucleic acid is in the aqueous interior. Multilamellar lipid nanoparticles can be complexed with nucleic acid, wherein the nucleic acid is in the aqueous interior, or to form or sandwiched between

[00189] Abbreviations: MHC: major histocompatibility complex; HLA: human leukocyte antigen, or the human MHC gene locus; NGS: next-generation sequencing; PPV: positive predictive value; TSNA: tumor-specific neoantigen; FFPE: formalin-fixed, paraffin-embedded; NMD: nonsense-mediated decay; NSCLC: non-small-cell lung cancer; DC: dendritic cell.

[00190] It should be noted that, as used in the specification and the appended claims, the singular forms “a,” “an,” and “the” include plural referents unless the context clearly dictates otherwise.

[00191] Unless specifically stated or otherwise apparent from context, as used herein the term “about” is understood as within a range of normal tolerance in the art, for example within 2 standard deviations of the mean. About can be understood as within 10%, 9%, 8%, 7%, 6%, 5%, 4%, 3%, 2%, 1%, 0.5%, 0.1%, 0.05%, or 0.01% of the stated value. Unless otherwise clear from context, all numerical values provided herein are modified by the term about.

[00192] Any terms not directly defined herein shall be understood to have the meanings commonly associated with them as understood within the art of the invention. Certain terms are discussed herein to provide additional guidance to the practitioner in describing the compositions, devices, methods and the like of aspects of the invention, and how to make or use them. It will be appreciated that the same thing may be said in more than one way.

Consequently, alternative language and synonyms may be used for any one or more of the terms discussed herein. No significance is to be placed upon whether or not a term is elaborated or discussed herein. Some synonyms or substitutable methods, materials and the like are provided. Recital of one or a few synonyms or equivalents does not exclude use of other synonyms or equivalents, unless it is explicitly stated. Use of examples, including examples of terms, is for illustrative purposes only and does not limit the scope and meaning of the aspects of the invention herein.

[00193] All references, issued patents and patent applications cited within the body of the specification are hereby incorporated by reference in their entirety, for all purposes.

II. Antigen Identification

[00194] 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, US App. No. 16/606,577, and international patent application publications WO2020181240A1, WO/2018/195357 and WO/2018/208856, each herein incorporated by reference, in their entirety, for all purposes.

[00195] 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 a tumor or an infectious disease organism); inputting the peptide sequence of each antigen into one or more presentation models to generate a set of numerical likelihoods that each of the antigens is presented by one or more MHC alleles on 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.

III. Identification of Tumor Specific Mutations in Neoantigens

[00196] Also disclosed herein are methods for the identification of certain mutations (*e.g.*, the variants or alleles that are present in cancer cells). In particular, these mutations can be present in the genome, transcriptome, proteome, or exome of cancer cells of a subject having cancer but not in normal tissue from the subject. Specific methods for identifying neoantigens, including shared neoantigens, that are specific to tumors 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. Examples of shared neoantigens that are specific to tumors are described in more detail in international patent application publication WO2019226941A1, herein incorporated by reference in its entirety, for all purposes. Shared neoantigens include, but are not limited to, KRAS-associated mutations (*e.g.*, KRAS G12C, KRAS G12V, KRAS G12D, and/or KRAS Q61H mutations). For example, KRAS-associated MHC class I neoepitope can include those mutations with reference to wild-type (WT) human KRAS, such as with reference to the the following exemplary amino acid sequence:

MTEYKLVVVGAGGVGKSALTIQLIQNHFVDEYDPTIEDSYRKQVVIDGETCLLDILDTA
GQEEYSAMRDQYMRTGEGFLCVFAINNTKSFEDIHHYREQIKRVKDSQEDVPMVLVGNK
CDLPSTVDTKQAQDLARSYGIPFIETSAKTRQRVEDAFYTLVREIRQYRLKKISKEEPTP
GCVKIKKCIIM (SEQ ID NO:84).

[00197] Genetic mutations in tumors can be considered useful for the immunological targeting of tumors if they lead to changes in the amino acid sequence of a protein exclusively in the tumor. Useful mutations include: (1) non-synonymous mutations leading to different amino acids in the protein; (2) read-through mutations in which a stop codon is modified or deleted, leading to translation of a longer protein with a novel tumor-specific sequence at the C-terminus; (3) splice site mutations that lead to the inclusion of an intron in the mature mRNA and thus a unique tumor-specific protein sequence; (4) chromosomal rearrangements that give rise to a chimeric protein with tumor-specific sequences at the junction of 2 proteins (i.e., gene fusion); (5) frameshift mutations or deletions that lead to a new open reading frame with a novel tumor-specific protein sequence. Mutations can also include one or more of non-frameshift indel, missense or nonsense substitution, splice site alteration, genomic rearrangement or gene fusion, or any genomic or expression alteration giving rise to a neoORF.

[00198] Peptides with mutations or mutated polypeptides arising from for example, splice-site, frameshift, readthrough, or gene fusion mutations in tumor cells can be identified by sequencing DNA, RNA or protein in tumor versus normal cells.

[00199] Also mutations can include previously identified tumor specific mutations. Known tumor mutations can be found at the Catalogue of Somatic Mutations in Cancer (COSMIC) database.

[00200] A variety of methods are available for detecting the presence of a particular mutation or allele in an individual's DNA or RNA. Advancements in this field have provided accurate, easy, and inexpensive large-scale SNP genotyping. For example, several techniques have been described including dynamic allele-specific hybridization (DASH), microplate array diagonal gel electrophoresis (MADGE), pyrosequencing, oligonucleotide-specific ligation, the TaqMan system as well as various DNA "chip" technologies such as the Affymetrix SNP chips. These methods utilize amplification of a target genetic region, typically by PCR. Still other methods, based on the generation of small signal molecules by invasive cleavage followed by mass spectrometry or immobilized padlock probes and rolling-circle amplification. Several of the methods known in the art for detecting specific mutations are summarized below.

[00201] PCR based detection means can include multiplex amplification of a plurality of markers simultaneously. For example, it is well known in the art to select PCR primers to generate PCR products that do not overlap in size and can be analyzed simultaneously. Alternatively, it is possible to amplify different markers with primers that are differentially labeled and thus can each be differentially detected. Of course, hybridization based detection means allow the differential detection of multiple PCR products in a sample. Other techniques are known in the art to allow multiplex analyses of a plurality of markers.

[00202] Several methods have been developed to facilitate analysis of single nucleotide polymorphisms in genomic DNA or cellular RNA. For example, a single base polymorphism can be detected by using a specialized exonuclease-resistant nucleotide, as disclosed, e.g., in Mundy, C. R. (U.S. Pat. No. 4,656,127). According to the method, a primer complementary to the allelic sequence immediately 3' to the polymorphic site is permitted to hybridize to a target molecule obtained from a particular animal or human. If the polymorphic site on the target molecule contains a nucleotide that is complementary to the particular exonuclease-resistant nucleotide derivative present, then that derivative will be incorporated onto the end of the hybridized primer. Such incorporation renders the primer resistant to exonuclease, and thereby permits its detection. Since the identity of the exonuclease-resistant derivative of the sample is known, a finding that the primer has become resistant to exonucleases reveals that the nucleotide(s) present in the polymorphic site of the target molecule is complementary to that of the nucleotide derivative used in the reaction. This method has the advantage that it does not require the determination of large amounts of extraneous sequence data.

[00203] A solution-based method can be used for determining the identity of a nucleotide of a polymorphic site. Cohen, D. et al. (French Patent 2,650,840; PCT Appln. No. WO91/02087). As in the Mundy method of U.S. Pat. No. 4,656,127, a primer is employed that is complementary to allelic sequences immediately 3' to a polymorphic site. The method determines the identity of the nucleotide of that site using labeled dideoxynucleotide derivatives, which, if complementary to the nucleotide of the polymorphic site will become incorporated onto the terminus of the primer.

[00204] An alternative method, known as Genetic Bit Analysis or GBA is described by Goelet, P. et al. (PCT Appln. No. 92/15712). The method of Goelet, P. et al. uses mixtures of labeled terminators and a primer that is complementary to the sequence 3' to a polymorphic site. The labeled terminator that is incorporated is thus determined by, and complementary to, the nucleotide present in the polymorphic site of the target molecule being evaluated. In contrast to the method of Cohen et al. (French Patent 2,650,840; PCT Appln. No. WO91/02087) the method of Goelet, P. et al. can be a heterogeneous phase assay, in which the primer or the target molecule is immobilized to a solid phase.

[00205] Several primer-guided nucleotide incorporation procedures for assaying polymorphic sites in DNA have been described (Komher, J. S. et al., Nucl. Acids. Res. 17:7779-7784 (1989); Sokolov, B. P., Nucl. Acids Res. 18:3671 (1990); Syvanen, A.-C., et al., Genomics 8:684-692 (1990); Kuppuswamy, M. N. et al., Proc. Natl. Acad. Sci. (U.S.A.) 88:1143-1147 (1991); Prezant, T. R. et al., Hum. Mutat. 1:159-164 (1992); Ugozzoli, L. et al., GATA 9:107-112 (1992); Nyren, P. et al., Anal. Biochem. 208:171-175 (1993)). These methods differ from GBA in that they utilize incorporation of labeled deoxynucleotides to discriminate between bases at a

polymorphic site. In such a format, since the signal is proportional to the number of deoxynucleotides incorporated, polymorphisms that occur in runs of the same nucleotide can result in signals that are proportional to the length of the run (Syvanen, A.-C., et al., *Amer. J. Hum. Genet.* 52:46-59 (1993)).

[00206] A number of initiatives obtain sequence information directly from millions of individual molecules of DNA or RNA in parallel. Real-time single molecule sequencing-by-synthesis technologies rely on the detection of fluorescent nucleotides as they are incorporated into a nascent strand of DNA that is complementary to the template being sequenced. In one method, oligonucleotides 30-50 bases in length are covalently anchored at the 5' end to glass cover slips. These anchored strands perform two functions. First, they act as capture sites for the target template strands if the templates are configured with capture tails complementary to the surface-bound oligonucleotides. They also act as primers for the template directed primer extension that forms the basis of the sequence reading. The capture primers function as a fixed position site for sequence determination using multiple cycles of synthesis, detection, and chemical cleavage of the dye-linker to remove the dye. Each cycle includes adding the polymerase/labeled nucleotide mixture, rinsing, imaging and cleavage of dye. In an alternative method, polymerase is modified with a fluorescent donor molecule and immobilized on a glass slide, while each nucleotide is color-coded with an acceptor fluorescent moiety attached to a gamma-phosphate. The system detects the interaction between a fluorescently-tagged polymerase and a fluorescently modified nucleotide as the nucleotide becomes incorporated into the de novo chain. Other sequencing-by-synthesis technologies also exist.

[00207] Any suitable sequencing-by-synthesis platform can be used to identify mutations. As described above, four major sequencing-by-synthesis platforms are currently available: the Genome Sequencers from Roche/454 Life Sciences, the 1G Analyzer from Illumina/Solexa, the SOLiD system from Applied BioSystems, and the Heliscope system from Helicos Biosciences. Sequencing-by-synthesis platforms have also been described by Pacific BioSciences and VisiGen Biotechnologies. In some embodiments, a plurality of nucleic acid molecules being sequenced is bound to a support (e.g., solid support). To immobilize the nucleic acid on a support, a capture sequence/universal priming site can be added at the 3' and/or 5' end of the template. The nucleic acids can be bound to the support by hybridizing the capture sequence to a complementary sequence covalently attached to the support. The capture sequence (also referred to as a universal capture sequence) is a nucleic acid sequence complementary to a sequence attached to a support that may dually serve as a universal primer.

[00208] As an alternative to a capture sequence, a member of a coupling pair (such as, e.g., antibody/antigen, receptor/ligand, or the avidin-biotin pair as described in, e.g., US Patent

Application No. 2006/0252077) can be linked to each fragment to be captured on a surface coated with a respective second member of that coupling pair.

[00209] Subsequent to the capture, the sequence can be analyzed, for example, by single molecule detection/sequencing, e.g., as described in the Examples and in U.S. Pat. No. 7,283,337, including template-dependent sequencing-by-synthesis. In sequencing-by-synthesis, the surface-bound molecule is exposed to a plurality of labeled nucleotide triphosphates in the presence of polymerase. The sequence of the template is determined by the order of labeled nucleotides incorporated into the 3' end of the growing chain. This can be done in real time or can be done in a step-and-repeat mode. For real-time analysis, different optical labels to each nucleotide can be incorporated and multiple lasers can be utilized for stimulation of incorporated nucleotides.

[00210] Sequencing can also include other massively parallel sequencing or next generation sequencing (NGS) techniques and platforms. Additional examples of massively parallel sequencing techniques and platforms are the Illumina HiSeq or MiSeq, Thermo PGM or Proton, the Pac Bio RS II or Sequel, Qiagen's Gene Reader, and the Oxford Nanopore MinION. Additional similar current massively parallel sequencing technologies can be used, as well as future generations of these technologies.

[00211] Any cell type or tissue can be utilized to obtain nucleic acid samples for use in methods described herein. For example, a DNA or RNA sample can be obtained from a tumor or a bodily fluid, e.g., blood, obtained by known techniques (e.g. venipuncture) or saliva. Alternatively, nucleic acid tests can be performed on dry samples (e.g. hair or skin). In addition, a sample can be obtained for sequencing from a tumor and another sample can be obtained from normal tissue for sequencing where the normal tissue is of the same tissue type as the tumor. A sample can be obtained for sequencing from a tumor and another sample can be obtained from normal tissue for sequencing where the normal tissue is of a distinct tissue type relative to the tumor.

[00212] Tumors can include one or more of lung cancer, melanoma, breast cancer, ovarian cancer, prostate cancer, kidney cancer, gastric cancer, colon cancer, testicular cancer, head and neck cancer, pancreatic cancer, brain cancer, B-cell lymphoma, acute myelogenous leukemia, chronic myelogenous leukemia, chronic lymphocytic leukemia, and T cell lymphocytic leukemia, non-small cell lung cancer, and small cell lung cancer.

[00213] Alternatively, protein mass spectrometry can be used to identify or validate the presence of mutated peptides bound to MHC proteins on tumor cells. Peptides can be acid-eluted from tumor cells or from HLA molecules that are immunoprecipitated from tumor, and then identified using mass spectrometry.

IV. Antigens

[00214] 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.

[00215] 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.

[00216] Specifically, disclosed herein cassettes including iterations of KRAS-associated MHC class I neoepitopes. KRAS-associated MHC class I neoepitopes include, but are not limited to, neoepitopes having KRAS G12 mutations and/or KRAS Q61 mutations. Cassettes can include iterations of KRAS-associated MHC class I neoepitopes having a KRAS G12 mutation. Cassettes can include iterations of KRAS-associated MHC class I neoepitopes having a KRAS Q61 mutation. Cassettes can include iterations of KRAS-associated MHC class I neoepitopes having KRAS G12C, KRAS G12V, KRAS G12D, and/or KRAS Q61H mutations. Cassettes can include iterations of KRAS-associated MHC class I neoepitopes having a KRAS G12C mutation. Cassettes can include iterations of KRAS-associated MHC class I neoepitopes having a KRAS G12V mutation. Cassettes can include iterations of KRAS-associated MHC class I neoepitopes having a KRAS G12D mutation. Cassettes can include iterations of KRAS-associated MHC class I neoepitopes having a KRAS Q61H mutation. Cassettes can include iterations of each of KRAS-associated MHC class I neoepitopes having a KRAS G12C, KRAS G12V, KRAS G12D, and KRAS Q61H mutation. Cassettes can include iterations of at least two distinct KRAS-associated MHC class I neoepitopes selected from the group consisting of: a KRAS G12C, KRAS G12V, KRAS G12D, and KRAS Q61H mutation. Cassettes can include iterations of at least three distinct KRAS-associated MHC class I neoepitopes selected from the group consisting of: a KRAS G12C, KRAS G12V, KRAS G12D, and KRAS Q61H mutation. Cassettes can include iterations only of a single distinct KRAS-associated MHC class I neoepitope. Cassettes can include iterations only of a single distinct KRAS-associated MHC class I neoepitope having a KRAS G12C mutation. Cassettes can include iterations only of a single distinct KRAS-associated MHC class I neoepitope having a KRAS G12D mutation. Cassettes can include iterations only of a single distinct KRAS-associated MHC class I neoepitope having a KRAS G12V mutation. Cassettes can include iterations only of a single distinct KRAS-associated MHC class I neoepitope having a KRAS Q61H mutation.

[00217] KRAS-associated MHC class I neoepitopes having a KRAS G12C mutation include VVVGACGVGK (SEQ ID NO: 75) or KLVVVGACGV (SEQ ID NO: 76). KRAS-associated MHC class I neoepitopes having a KRAS G12D mutation include VVGADGVGK (SEQ ID NO: 77) or VVVGADGVGK (SEQ ID NO: 78), KRAS-associated MHC class I neoepitopes having a KRAS G12V mutation include VVGAVGVGK (SEQ ID NO: 79), VVVGAVGVGK (SEQ ID NO: 81), or AVGVGKSAL (SEQ ID NO: 80).

[00218] Cassettes can include iterations of each of KRAS-associated MHC class I neoepitopes having the amino acid sequences VVVGACGVGK (SEQ ID NO: 75), VVVGADGVGK (SEQ ID NO: 78), VVGAVGVGK (SEQ ID NO: 79), and ILDTAGHEEY (SEQ ID NO: 82).

Cassettes can include iterations of at least two distinct KRAS-associated MHC class I neoepitopes having the amino acid sequences selected from the group consisting of:

VVVGACGVGK (SEQ ID NO: 75), VVVGADGVGK (SEQ ID NO: 78), VVGAVGVGK (SEQ ID NO: 79), and ILDTAGHEEY (SEQ ID NO: 82). Cassettes can include iterations of at least three distinct KRAS-associated MHC class I neoepitopes having the amino acid sequences selected from the group consisting of: VVVGACGVGK (SEQ ID NO: 75), VVVGADGVGK (SEQ ID NO: 78), VVGAVGVGK (SEQ ID NO: 79), and ILDTAGHEEY (SEQ ID NO: 82).

Cassettes can include iterations of at least one of KRAS-associated MHC class I neoepitopes having the amino acid sequences VVVGACGVGK (SEQ ID NO: 75), VVVGADGVGK (SEQ ID NO: 78), VVGAVGVGK (SEQ ID NO: 79), and ILDTAGHEEY (SEQ ID NO: 82).

[00219] KRAS-associated MHC class I neoepitopes can include native N- and/or C-terminal flanking sequences of the therapeutic vaccine epitope in the context of the native KRAS protein. Illustrative non-limiting examples of KRAS-associated MHC class I neoepitopes are the 25mers MTEYKLVVVGACGVGKSALTIQLIQ (SEQ ID NO: 57) for KRAS G12C, MTEYKLVVVGADGVGKSALTIQLIQ (SEQ ID NO: 58) for KRAS G12D, MTEYKLVVVGAVGVGKSALTIQLIQ (SEQ ID NO: 59) for KRAS G12V, and ETCLLDILDTAGHEEYSAMRDQYMR (SEQ ID NO: 60) for KRAS Q61H. KRAS-associated MHC class I neoepitopes that include native flanking sequences can be linked (concatenated) to other neoepitopes encoded in a cassette, including other neoepitopes (*e.g.*, other KRAS-associated MHC class I neoepitopes) that include their respective native flanking sequences. An illustrative non-limiting cassette of concatenated KRAS-associated MHC class I neoepitopes that are linked through their native flanking sequences and that includes 4 iterations for each of the KRAS neoepitopes having the mutations KRAS G12C, KRAS G12D, KRAS G12V, and KRAS Q61H is represented by the amino acid sequence shown in SEQ ID NO: 65.

[00220] Epitope-encoding nucleic acid sequences that encode KRAS-associated MHC class I neoepitopes, such as those that include native N- and/or C-terminal flanking sequences, can

encode multiple known and/or predicted KRAS-associated MHC class I neoepitopes. As an illustrative example, the KRAS G12V 25mer MTEYKLVVVGAVGVGKSALTIQLIQ (SEQ ID NO: 59) encodes each of the known and/or predicted KRAS-associated MHC class I neoepitopes VVGAVGVGK (SEQ ID NO: 79), VVVGAVGVGK (SEQ ID NO: 81), and AVGVGKSAL (SEQ ID NO: 80).

[00221] Epitope-encoding nucleic acid sequences, including those that encode KRAS-associated MHC class I neoepitopes, can be in any order in a cassette. Epitope-encoding nucleic acid sequences, including those that encode KRAS-associated MHC class I neoepitopes, can be in an order that minimizes junctional epitopes, as described further herein. As an illustrative non-limiting example, concatenated KRAS-associated MHC class I neoepitopes linked together to minimize junctional epitopes is represented by the amino acid sequence shown in SEQ ID NO: 65 and has the order: G12C G12D Q61H G12D G12V G12C Q61H G12D G12V G12C Q61H G12D G12V Q61H G12V G12C.

[00222] 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. The peptide contains the 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. 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.

[00223] 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 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, an orthomyxoviridae family virus, and tuberculosis.

[00224] 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.

[00225] 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.

[00226] 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 hantavirus fever virus), nairovirus (Nairobi sheep disease) and various unassigned bunyaviruses. The arenavirus family provides a source of antigens against LCM and Lassa fever virus. The reovirus family includes the genera reovirus, rotavirus (which causes acute gastroenteritis in children), orbiviruses, and cultivirus (Colorado Tick fever, Lebombo (humans), equine encephalosis, blue tongue). The retrovirus family includes the sub-family oncovirinae which encompasses such human and veterinary diseases as feline leukemia virus, HTLV-I and HTLV-II, lentivirinae (which includes human immunodeficiency virus (HIV), simian immunodeficiency virus (SIV), feline immunodeficiency virus (FIV), equine infectious anemia virus, and spumavirinae). Among the lentiviruses, many suitable antigens have been described and can readily be selected. Examples of suitable HIV and SIV antigens include, without limitation the gag, pol, Vif, Vpx, VPR, Env, Tat, Nef, and Rev proteins, as well as various fragments thereof. For example, suitable fragments of the Env protein may include any of its subunits such as the gp120, gp160, gp41, or smaller fragments thereof, e.g., of at least about 8 amino acids in length. Similarly, fragments of the tat protein may be selected. [See, U.S. Pat. No. 5,891,994 and U.S. Pat. No. 6,193,981.] See, also, the HIV and SIV proteins described in D. H. Barouch et al, *J. Virol.*, 75(5):2462-2467 (March 2001), and R. R. Amara, et al, *Science*, 292:69-74 (6 Apr. 2001). In another example, the HIV and/or SIV immunogenic proteins or peptides may be used to form fusion proteins or other immunogenic molecules. See, e.g., the HIV-1 Tat and/or Nef fusion proteins and immunization regimens described in WO 01/54719, published Aug. 2, 2001, and WO 99/16884, published Apr. 8, 1999. The invention is not limited to the HIV and/or SIV immunogenic proteins or peptides described herein. In addition, a variety of modifications to these proteins have been described or could readily be made by one of skill in the art. See, e.g., the modified gag protein that is described in U.S. Pat. No. 5,972,596. Further, any desired HIV and/or SIV immunogens may be delivered alone or in combination. Such

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

[00227] Antigens that can be incorporated into a vaccine (*e.g.*, encoded in a cassette) also include immunogens which are useful to immunize a human or non-human animal against pathogens including bacteria, fungi, parasitic microorganisms or multicellular parasites which infect human and non-human vertebrates. Examples of bacterial pathogens include pathogenic gram-positive cocci include pneumococci; staphylococci; and streptococci. Pathogenic gram-negative cocci include meningococcus; gonococcus. Pathogenic enteric gram-negative bacilli include enterobacteriaceae; *pseudomonas*, *acinetobacteria* and *eikenella*; melioidosis; *salmonella*; *shigella*; *haemophilus* (*Haemophilus influenzae*, *Haemophilus somnus*); *moraxella*; *H. ducreyi* (which causes chancroid); *brucella*; *Francisella tularensis* (which causes tularemia); *Yersinia (pasteurella)*; *Streptobacillus moniliformis* and *Spirillum*. Gram-positive bacilli include *Listeria monocytogenes*; *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.

[00228] 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, an orthomyxoviridae family virus, and tuberculosis.

[00229] Antigens can be selected that are predeicted 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.

[00230] 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.

[00231] 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.

[00232] 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.

[00233] 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 a tumor or 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 a tumor or an infectious disease antigen can be an antigen found on the surface of tumor cell or an infectious disease organism, respectively. Antigens capable of eliciting a B cell response to a tumor or an infectious disease antigen can be an intracellular neoantigen expressed in a tumor or an infectious disease organism, respectively.

[00234] 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).

[00235] 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.

[00236] 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.

[00237] 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.

[00238] 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 a tumor or an infectious disease organism, respectively. Longer peptides (*e.g.*, full-length protein, protein subunit, or protein domain) and combinations thereof can be included to stimulate a B cell response.

[00239] 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.

[00240] In some aspects, antigenic peptides and polypeptides do not stimulate an autoimmune response and/or invoke immunological tolerance when administered to a subject.

[00241] Also provided are compositions comprising at least two or more antigenic peptides. In some embodiments the composition contains at least two distinct peptides. At least two distinct peptides can be derived from the same polypeptide. By distinct polypeptides is meant that the peptide vary by length, amino acid sequence, or both. A peptide can include a tumor-specific mutation. 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. The peptides can be derived from any polypeptide known to or suspected to be associated with an infectious disease organism, or peptides derived from any polypeptide known to or have been found to have altered expression in an infected cell in comparison to a normal cell or tissue (*e.g.*, an infectious disease polynucleotide or polypeptide, including infectious disease polynucleotides or polypeptides with expression restricted to a host cell). 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. In some aspects the tumor specific mutation is a driver mutation for a particular cancer type. A peptide can include a KRAS mutation (*e.g.*, KRAS G12C, KRAS G12V, KRAS G12D, and/or KRAS Q61H mutations).

[00242] 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).

[00243] 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.

[00244] 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.

[00245] 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.

[00246] 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.

[00247] 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 phosphorothioate 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.

[00248] 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.

V. Vaccine Compositions

[00249] 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.

[00250] 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.

[00251] 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).

[00252] 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.

[00253] A vaccine can contain at least two iterations of an epitope-encoding nucleic acid sequence. A used herein, an “iteration” (or interchangeably a “repeat”) refers to two or more identical nucleic acid epitope-encoding nucleic acid sequences (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 iterations 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 iterations 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 iterations of at least one of the distinct epitopes are illustrated by, but is not limited to, the formulas below:

- Iteration of one distinct epitope (iteration of epitope A):

E_A-E_B-E_C-E_A; or

E_A-E_A-E_B-E_C

- Iteration of multiple distinct epitopes (iterations 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 iterations of multiple distinct epitopes (iterations 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

[00254] The above examples are not limiting and the antigen-encoding nucleic acid sequences having iterations 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.

[00255] Also provided for herein is an antigen-encoding cassette, the antigen-encoding cassette having at least one antigen-encoding nucleic acid sequence described, from 5’ to 3’, by the formula:

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

where E represents a nucleotide sequence including a 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. In some aspects, at least one of the distinct epitope-encoding nucleic acid sequences with the at least two iterations encodes a KRAS-associated MHC class I neoepitope.

[00256] 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.

[00257] Iterations 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). Iterations of an epitope-encoding nucleic acid sequences can be separated by one or more additional nucleotides sequences. In general, iterations of an epitope-encoding nucleic acid sequences can be separated by any size nucleotide sequence applicable for the compositions described herein. In one example, iterations 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 iterations 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 iterations 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

VTNTEMFVTAPDNLGYMYEVQWPGQTQPQIANCSVYDFFVWLHYYSVRDTVTNTEMF
VTAPDNLGYMYEVQWPGQTQPQIANCSVYDFFVWLHYYSVRDT encoding iterations of
25mer antigens Trp1 (VTNTEMFVTAPDNLGYMYEVQWPGQ) and Trp2

(TQPQIANCSVYDFFVWLHYYSVRDT), the iterations 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 iterations 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 iterations can be separated by 150, 225, 300, 375, 450, 525, 600, or 675 nucleotides, respectively.

[00258] 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.

[00259] 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.

[00260] The vaccine composition can be capable of stimulating a specific B-cell response (e.g., an antibody response).

[00261] 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.

[00262] 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.

[00263] 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.

[00264] 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.

[00265] Suitable adjuvants include, but are not limited to 1018 ISS, alum, aluminum 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).

[00266] 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.

[00267] 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).

[00268] 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.

[00269] 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.

[00270] 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.

[00271] 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.

V.A. Antigen Cassette

[00272] 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” or “cassette” is meant the combination of a selected antigen or plurality of antigens (e.g., antigen-encoding nucleic acid sequences) and the other regulatory elements necessary to transcribe the antigen(s) and express the transcribed product. The selected antigen or plurality of antigens can refer to distinct epitope sequences, e.g., an antigen-encoding nucleic acid sequence in the cassette can encode an epitope-encoding nucleic acid sequence (or plurality of epitope-encoding nucleic acid sequences) such that the epitopes are transcribed and expressed. An antigen or plurality of antigens can be operatively linked to regulatory components in a manner which permits transcription. Such components include conventional regulatory elements that can drive expression of the antigen(s) in a cell transfected with the viral vector. Thus the antigen cassette can also contain a selected promoter which is linked to the antigen(s) and located, with other, optional regulatory elements, within the selected viral sequences of the recombinant vector. 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).

[00273] 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.

[00274] Also disclosed herein is a viral vector comprising a cassette with at least one payload sequence operably linked to a regulatable promoter that is a TET promoter system, such as a TET-On system or TET-Off system. Without wishing to be bound by theory, a TET promoter system can be used to minimize transcription of payload nucleic acids encoded in a cassette, such as antigens encoded in a vaccine cassette, during viral production. TET promoter systems are described in detail in international patent application publication WO2020/243719, herein incorporated by reference for all purposes.

[00275] A TET promoter system can include a tetracycline (TET) repressor protein (TET_r) controlled promoter. Accordingly, also disclosed herein is a viral vector comprising a cassette with at least one payload sequence operably linked to a tetracycline (TET) repressor protein (TET_r) controlled promoter. A TET_r controlled promoter can include the 19 bp TET operator (TET_o) sequence TCCCTATCAGTGATAGAGA (SEQ ID NO:83). A TET_r controlled promoter can include 2, 3, 4, 5, 6, 7, 8, 9, or 10 or more TET_o nucleic acid sequences. In TET_r controlled promoter have 2 or more TET_o nucleic acid sequences, the TET_o sequences can be linked together. In TET_r controlled promoter have 2 or more TET_o nucleic acid sequences, the TET_o sequences can be directly linked together. In TET_r controlled promoter have 2 or more TET_o nucleic acid sequences, the TET_o sequences can be linked together with a linker sequence,

such as a linker sequence having 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 or more nucleotides. In general, a TETr controlled promoter can use any promoter sequence desired, such as a SV40, EF-1, RSV, PGK, HSA, MCK or EBV promoter sequence. A TETr controlled promoter can use a CMV promoter sequence. A TETr controlled promoter can use a minimal CMV promoter sequence. TETo sequences can be upstream (5') of a promoter sequence region where RNA polymerase binds. In an illustrative example, 7 TETo sequences are upstream (5') of a promoter sequence. A TETr controlled promoter operably linked to the at least one payload nucleic acid sequence with TETo sequence upstream of the promoter sequence region can have an ordered sequence described in the formula, from 5' to 3':

$$(T-L_Y)_X-P-N$$

where N is a payload nucleic acid sequence, P is a RNA polymerase binding sequence of the promoter sequence operably linked to payload nucleic acid sequence, T is a TETo nucleic acid sequences comprising the nucleotide sequence shown in SEQ ID NO:66, L is a linker sequence, where Y = 0 or 1 for each X, and wherein X = 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20. In an illustrative example, X = 7 and Y = 1 for each X describes where 7 TETo sequences are upstream (5') of the promoter sequence and each TETo sequence is separated by a linker.

[00276] A TETo sequences can be downstream (3') of a promoter sequence region where RNA polymerase binds. In another illustrative example, 2 TETo sequences are downstream (3') of a promoter sequence. A TETr controlled promoter operably linked to the at least one payload nucleic acid sequence with TETo sequence downstream of the promoter sequence region can have an ordered sequence described in the formula, from 5' to 3':

$$P-(T-L_Y)_X-N$$

where N is a payload nucleic acid sequence, P is a RNA polymerase binding sequence of the promoter sequence operably linked to payload nucleic acid sequence, T is a TETo nucleic acid sequences comprising the nucleotide sequence shown in SEQ ID NO:66, L is a linker sequence, where Y = 0 or 1 for each X, and wherein X = 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20. In an illustrative example, X = 2 and Y = 1 for each X describes where 2 TETo sequences are downstream (3') of the promoter sequence and each TETo sequence is separated by a linker.

[00277] Viral production of vectors with TETr controlled promoters can use any viral production cell line engineered to express a TETr sequence (tTS), such as a 293 cell line or its derivatives (*e.g.*, a 293F cell line) engineered to express tTS. Viral production of vectors with TETr controlled promoters in tTS-expressing cell can improve viral production. Viral production of vectors with TETr controlled promoters in tTS-expressing cell can improve viral infectivity

defined as viral particles (VP) per infectious unit (IU). Viral production of vectors with TET_r controlled promoters in tTS-expressing cell can improve viral production and/or viral infectivity by at least 1.5, at least 2, at least 2.5, at least 3, at least 3.5, at least 4, at least 4.5, at least 5, at least 6, at least 7, at least 8, at least 9, or at least 10-fold relative to production in a non-tTS-expressing cell. Viral production of vectors with TET_r controlled promoters in tTS-expressing cell can improve viral production and/or viral infectivity by 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 70, at least 80, at least 90, or at least 100-fold relative to production in a non-tTS-expressing cell. Viral production of vectors with TET_r controlled promoters in tTS-expressing cell can improve viral production and/or viral infectivity by at least 1.5, at least 2, at least 2.5, at least 3, at least 3.5, at least 4, at least 4.5, at least 5, at least 6, at least 7, at least 8, at least 9, or at least 10-fold relative to production of a vector not having a TET_r controlled promoter. Viral production of vectors with TET_r controlled promoters in tTS-expressing cell can improve viral production and/or viral infectivity by 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 70, at least 80, at least 90, or at least 100-fold relative to production of a vector not having a TET_r controlled promoter.

[00278] The antigen cassette can also include nucleic acid sequences heterologous to the viral vector sequences including sequences providing signals for efficient polyadenylation of the transcript (poly(A), poly-A or pA) and introns with functional splice donor and acceptor sites. A common poly-A sequence which is employed in the exemplary vectors of this invention is that derived from the papovavirus SV-40. The poly-A sequence generally can be inserted in the cassette following the antigen-based sequences and before the viral vector sequences. A common intron sequence can also be derived from SV-40, and is referred to as the SV-40 T intron sequence. An antigen cassette can also contain such an intron, located between the promoter/enhancer sequence and the antigen(s). Selection of these and other common vector elements are conventional [see, e.g., Sambrook et al, "Molecular Cloning. A Laboratory Manual.", 2d edit., Cold Spring Harbor Laboratory, New York (1989) and references cited therein] and many such sequences are available from commercial and industrial sources as well as from Genbank.

[00279] 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.

[00280] 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 VEE 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.

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

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

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 an epitope encoding nucleic acid sequence, where for each Y the corresponding U_f is a MHC class II epitope-encoding nucleic acid sequence (*e.g.*, 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.

[00282] 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 a subgenomic promoter sequence (*e.g.*, a 26S subgenomic promoter sequence), an alphavirus 5' UTR, a 51-nt CSE, or a 24-nt CSE.

[00283] 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).

[00284] 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.

[00285] 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.

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

[00287] 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 sequence) 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.

[00288] The 5' linker L5 can be a native sequence or a non-natural sequence. Non-natural sequence include, but are not limited to, AAY, RR, and DPP. The 3' linker L3 can also be a native sequence or a non-natural sequence. Additionally, L5 and L3 can both be native sequences, both be non-natural sequences, or one can be native and the other non-natural. For each X, the amino acid linkers can be 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100 or more amino acids in length. For each X, the amino acid linkers can be also be at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 11, at least 12, at least 13, at least 14, at least 15, at least 16, at least 17, at least 18, at least 19, at least 20, at least 21, at least 22, at least 23, at least 24, at least 25, at least 26, at least 27, at least 28, at least 29, or at least 30 amino acids in length.

[00289] The amino acid linker G5, for each Y, can be 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100 or more amino acids in length. For each Y, the amino acid linkers can be also be at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 11, at least 12, at least 13, at least 14, at least 15, at least 16, at least 17, at least 18, at least 19, at least 20, at least 21, at least 22, at least 23, at least 24, at least 25, at least 26, at least 27, at least 28, at least 29, or at least 30 amino acids in length.

[00290] The amino acid linker G3 can be 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100 or more amino acids in length. G3 can be also be at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 11, at least 12, at least 13, at least 14, at least 15, at least 16, at least 17, at least 18, at least 19, at least 20, at least 21, at least 22, at least 23, at least 24, at least 25, at least 26, at least 27, at least 28, at least 29, or at least 30 amino acids in length.

[00291] 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 a MHC class I epitope 7-15 amino acids in length. For each X, each N can also encode a MHC class I epitope 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, or 30 amino acids in length. For each X, each N can also encode a MHC class I epitope at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 11, at least 12, at least 13, at least 14, at least 15, at least 16, at least 17, at least 18, at least 19, at least 20, at least 21, at least 22, at least 23, at least 24, at least 25, at least 26, at least 27, at least 28, at least 29, or at least 30 amino acids in length.

[00292] 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.

[00293] 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 (*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 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 can 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.

[00294] 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.

[00295] 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.

[00296] In some instances, an antigen or epitope in a cassette encoding additional antigens and/or epitopes may be an immunodominant epitope relative to the others encoded. Immunodominance, in general, is the skewing of an immune response towards only one or a few specific immunogenic peptides. Immunodominance can be assessed as part of an immune monitoring protocol. For example, immunodominance can be assessed through evaluating T cell and/or B cell responses to the encoded antigens.

[00297] Immunodominance can be assessed as the impact of an immunodominant antigen's presence on the immune response to one or more other antigens. For example, an immunodominant antigen and its respective immune response (*e.g.*, an immunodominant MHC class I epitope) can reduce the immune response of another antigen relative to the immune response in the absence of the immunodominant antigen. This reduction can be such that the immune response in the presence of the immunodominant antigen is not considered a

therapeutically effective response. For example, an MHC class I epitope would generally be considered immunodominant if T cell responses to other antigens are no longer considered therapeutically effective responses compared to responses elicited in the absence of the immunodominant MHC class I epitope. An immune response can also be reduced to below a limit of detection or near the limit of detection, relative to the response in the absence of the immunodominant antigen. For example, an MHC class I epitope would generally be considered immunodominant if T cell responses to other antigens are at or below the limit of detection compared to responses elicited in the absence of the immunodominant MHC class I epitope. In general, the assessment of immunodominance is between two antigens both capable of stimulating an immune response, *e.g.*, between two T cell epitopes in a vaccine composition administered to a subject possessing a cognate MHC allele known or predicted to present each epitope, respectively. Immunodominance can be assessed through evaluating relative immune responses to other antigens in the presence and absence of the suspected immunodominant antigen.

[00298] Immunodominance can be assessed as a relative difference in the immune responses between two or more antigens. Immunodominance can refer to a 5-fold, 10-fold, 20-fold, 30-fold, 40-fold, or 50-fold immune response of a specific antigen relative to another antigen encoded in the same cassette. Immunodominance can refer to a 100-fold, 200-fold, 300-fold, 400-fold, or 500-fold immune response of a specific antigen relative to another antigen encoded in the same cassette. Immunodominance can refer to a 1000-fold, 2000-fold, 3000-fold, 4000-fold, or 5000-fold immune response of a specific antigen relative to another antigen encoded in the same cassette. Immunodominance can refer to a 10,000-fold immune response of a specific antigen relative to another antigen encoded in the same cassette.

[00299] In some instances, it may be desired to avoid vaccine compositions containing an immunodominant epitope. For example, it may be desired to avoid designing a vaccine cassette encoding an immunodominant epitope. Without wishing to be bound by theory, administering and/or encoding an immunodominant epitope together with additional epitope may reduce the immune response to the additional epitopes, including potentially ultimately reducing vaccine efficacy against the additional epitopes. As an illustrative non-limiting example, vaccine compositions including TP53-associated neoepitopes may have the immune response, *e.g.*, a T cell response, skewed towards the TP53-associated neoepitope negatively impacting (*e.g.*, reducing the immune response to where the immune response is not a therapeutically effective response and/or to below a limit of detection) the immune response to other antigens or epitopes in the vaccine composition (*e.g.*, one or more KRAS-associated neoepitopes in the vaccine composition, such as any of the KRAS-associated neoepitopes shown in SEQ ID NOs. 75-82).

Accordingly, vaccine compositions can be designed to not contain an immunodominant epitope, such as designing a vaccine cassette (*e.g.*, a (neo)antigen-encoding cassette) to not encode an immunodominant epitope. For example, the cassette does not encode an epitope that reduces an immune response to another epitope encoded in the cassette when administered in a vaccine composition to a subject relative to an immune response when the other epitope is administered in the absence of the immunodominant MHC class I epitope. In another example, the cassette does not encode an epitope that reduces an immune response to another epitope encoded in the cassette to below a limit of detection when administered in a vaccine composition to a subject relative to an immune response when the other epitope is administered in the absence of the immunodominant MHC class I epitope. In another example, the cassette does not encode an epitope that reduces an immune response to another epitope encoded in the cassette, wherein the immune response is not a therapeutically effective response, when administered in a vaccine composition to a subject relative to an immune response when the other epitope is administered in the absence of the immunodominant MHC class I epitope. In another example, the cassette does not encode an epitope that stimulates a 5-fold, 10-fold, 20-fold, 30-fold, 40-fold, or 50-fold or greater immune response relative to another epitope encoded in the same cassette in a vaccine composition administered to a subject, where each antigen is capable of stimulating an immune response in the subject. In another example, the cassette does not encode an epitope that stimulates a 100-fold, 200-fold, 300-fold, 400-fold, or 500-fold or greater immune response relative to another epitope encoded in the same cassette in a vaccine composition administered to a subject, where each antigen is capable of stimulating an immune response in the subject. In another example, the cassette does not encode an epitope that stimulates a 1000-fold, 2000-fold, 3000-fold, 4000-fold, or 5000-fold or greater immune response relative to another epitope encoded in the same cassette in a vaccine composition administered to a subject, where each antigen is capable of stimulating an immune response in the subject. In another example, the cassette does not encode an epitope that results in a 10,000-fold or greater immune response relative to another epitope encoded in the same cassette in a vaccine composition administered to a subject, where each antigen is capable of stimulating an immune response in the subject.

V.B. Immune Modulators

[00300] 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.

[00301] 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.

V.C. Additional Considerations for Vaccine Design and Manufacture

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

Subclones

[00302] Truncal peptides, meaning those presented by all or most tumor 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 tumor subclones and choosing peptides so as to maximize the number of tumor subclones covered by a vaccine.

V.C.2. Antigen Prioritization

[00303] 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)

[00304] 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.

V.D. Alphavirus

V.D.1. Alphavirus Biology

[00305] 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).

[00306] 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.

[00307] 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.

V.D.2. Alphavirus as a delivery vector

[00308] 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 VEE. Thus, an example of an antigen expression vector described herein can utilize an alphavirus backbone that allows for a high level of antigen expression, 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 VEE or its attenuated derivative TC-83.

[00309] Several expression vector design strategies have been engineered using alphavirus sequences (Pushko 1997). In one strategy, a alphavirus vector design includes inserting a second copy of the 26S promoter sequence elements downstream of the structural protein genes, followed by a heterologous gene (Frolov 1993). Thus, in addition to the natural non-structural and structural proteins, an additional subgenomic RNA is produced that expresses the heterologous protein. In this system, all the elements for production of infectious virions are present and, therefore, repeated rounds of infection of the expression vector in non-infected cells can occur.

[00310] Another expression vector design makes use of helper virus systems (Pushko 1997). In this strategy, the structural proteins are replaced by a heterologous gene. Thus, following self-replication of viral RNA mediated by still intact non-structural genes, the 26S subgenomic RNA provides for expression of the heterologous protein. Traditionally, additional vectors that expresses the structural proteins are then supplied *in trans*, such as by co-transfection of a cell line, to produce infectious virus. A system is described in detail in USPN 8,093,021, which is herein incorporated by reference in its entirety, for all purposes. The helper vector system provides the benefit of limiting the possibility of forming infectious particles and, therefore, improves biosafety. In addition, the helper vector system reduces the total vector length, potentially improving the replication and expression efficiency. Thus, an example of an antigen expression vector described herein can utilize an alphavirus backbone wherein the structural

proteins are replaced by an antigen cassette, the resulting vector both reducing biosafety concerns, while at the same time promoting efficient expression due to the reduction in overall expression vector size.

V.D.3. Alphavirus production *in vitro*

[00311] Alphavirus delivery vectors are generally positive-sense RNA polynucleotides. A convenient technique well-known in the art for RNA production is *in vitro* transcription IVT. In this technique, a DNA template of the desired vector is first produced by techniques well-known to those in the art, including standard molecular biology techniques such as cloning, restriction digestion, ligation, gene synthesis (*e.g.*, chemical and/or enzymatic synthesis), and polymerase chain reaction (PCR). The DNA template contains a RNA polymerase promoter at the 5' end of the sequence desired to be transcribed into RNA. Promoters include, but are not limited to, bacteriophage polymerase promoters such as T3, T7, or SP6. The DNA template is then incubated with the appropriate RNA polymerase enzyme, buffer agents, and nucleotides (NTPs). 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. The RNA can then be purified using techniques well-known in the field, such as phenol-chloroform extraction or column purification (*e.g.*, chromatography-based purification).

V.D.4. Delivery via lipid nanoparticle

[00312] 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.

[00313] 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.

[00314] 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.

[00315] Lipid nanoparticles (LNPs) are an attractive delivery system due to the amphiphilic nature of lipids enabling formation of membranes and vesicle like structures (Riley 2017). In general, these vesicles deliver the expression vector by absorbing into the membrane of target cells and releasing nucleic acid into the cytosol. In addition, LNPs can be further modified or functionalized to facilitate targeting of specific cell types. Another consideration in LNP design is the balance between targeting efficiency and cytotoxicity. Lipid compositions generally include defined mixtures of cationic, neutral, anionic, and amphipathic lipids. In some instances, specific lipids are included to prevent LNP aggregation, prevent lipid oxidation, or provide functional chemical groups that facilitate attachment of additional moieties. Lipid composition can influence overall LNP size and stability. In an example, the lipid composition comprises dilinoleylmethyl- 4-dimethylaminobutyrate (MC3) or MC3-like molecules. MC3 and MC3-like lipid compositions can be formulated to include one or more other lipids, such as a PEG or PEG-conjugated lipid, a sterol, or neutral lipids.

[00316] 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 affects. 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.

V.E. Chimpanzee adenovirus (ChAd)

V.E.1. Viral delivery with chimpanzee adenovirus

[00317] Vaccine compositions for delivery of one or more antigens (e.g., via an antigen cassette) can be created by providing adenovirus nucleotide sequences of chimpanzee origin, a variety of novel vectors, and cell lines expressing chimpanzee adenovirus genes. A nucleotide sequence of a chimpanzee C68 adenovirus (also referred to herein as ChAdV68) can be used in a vaccine composition for antigen delivery (*See* SEQ ID NO: 1). Use of C68 adenovirus derived vectors is described in further detail in USPN 6,083,716, which is herein incorporated by reference in its entirety, for all purposes. ChAdV68-based vectors and delivery systems are described in detail in US App. Pub. No. US20200197500A1 and international patent application publication WO2020243719A1, each of which is herein incorporated by reference for all purposes.

[00318] In a further aspect, provided herein is a recombinant adenovirus comprising the DNA sequence of a chimpanzee adenovirus such as C68 and an antigen cassette operatively linked to regulatory sequences directing its expression. The recombinant virus is capable of infecting a mammalian, preferably a human, cell and capable of expressing the antigen cassette product in the cell. In this vector, the native chimpanzee E1 gene, and/or E3 gene, and/or E4 gene can be deleted. An antigen cassette can be inserted into any of these sites of gene deletion. The antigen cassette can include an antigen against which a primed immune response is desired.

[00319] In another aspect, provided herein is a mammalian cell infected with a chimpanzee adenovirus such as C68.

[00320] In still a further aspect, a novel mammalian cell line is provided which expresses a chimpanzee adenovirus gene (e.g., from C68) or functional fragment thereof.

[00321] In still a further aspect, provided herein is a method for delivering an antigen cassette into a mammalian cell comprising the step of introducing into the cell an effective amount of a chimpanzee adenovirus, such as C68, that has been engineered to express the antigen cassette.

[00322] Still another aspect provides a method for stimulating an immune response in a mammalian host to treat cancer. The method can comprise the step of administering to the host an effective amount of a recombinant chimpanzee adenovirus, such as C68, comprising an antigen cassette that encodes one or more antigens from the tumor against which the immune response is targeted.

[00323] Still another aspect provides a method for stimulating an immune response in a mammalian host to treat or prevent a disease in a subject, such as an infectious disease. The method can comprise the step of administering to the host an effective amount of a recombinant chimpanzee adenovirus, such as C68, comprising an antigen cassette that encodes one or more antigens, such as from the infectious disease against which the immune response is targeted.

[00324] Also disclosed is a non-simian mammalian cell that expresses a chimpanzee adenovirus gene obtained from the sequence of SEQ ID NO: 1. The gene can be selected from the group consisting of the adenovirus E1A, E1B, E2A, E2B, E3, E4, L1, L2, L3, L4 and L5 of SEQ ID NO: 1.

[00325] Also disclosed is a nucleic acid molecule comprising a chimpanzee adenovirus DNA sequence comprising a gene obtained from the sequence of SEQ ID NO: 1. The gene can be selected from the group consisting of said chimpanzee adenovirus E1A, E1B, E2A, E2B, E3, E4, L1, L2, L3, L4 and L5 genes of SEQ ID NO: 1. In some aspects the nucleic acid molecule comprises SEQ ID NO: 1. In some aspects the nucleic acid molecule comprises the sequence of SEQ ID NO: 1, lacking at least one gene selected from the group consisting of E1A, E1B, E2A, E2B, E3, E4, L1, L2, L3, L4 and L5 genes of SEQ ID NO: 1.

[00326] Also disclosed is a vector comprising a chimpanzee adenovirus DNA sequence obtained from SEQ ID NO: 1 and an antigen cassette operatively linked to one or more regulatory sequences which direct expression of the cassette in a heterologous host cell, optionally wherein the chimpanzee adenovirus DNA sequence comprises at least the *cis*-elements necessary for replication and virion encapsidation, the *cis*-elements flanking the antigen cassette and regulatory sequences. In some aspects, the chimpanzee adenovirus DNA sequence comprises a gene selected from the group consisting of E1A, E1B, E2A, E2B, E3, E4, L1, L2, L3, L4 and L5 gene sequences of SEQ ID NO: 1. In some aspects the vector can lack the E1A and/or E1B gene.

[00327] Also disclosed herein is an adenovirus vector comprising: a partially deleted E4 gene comprising a deleted or partially-deleted E4orf2 region and a deleted or partially-deleted E4orf3 region, and optionally a deleted or partially-deleted E4orf4 region. The partially deleted E4 can comprise an E4 deletion of at least nucleotides 34,916 to 35,642 of the sequence shown in SEQ ID NO:1, and wherein the vector comprises at least nucleotides 2 to 36,518 of the sequence set forth in SEQ ID NO:1. The partially deleted E4 can comprise an E4 deletion of at least a partial deletion of nucleotides 34,916 to 34,942 of the sequence shown in SEQ ID NO:1, at least a partial deletion of nucleotides 34,952 to 35,305 of the sequence shown in SEQ ID NO:1, and at least a partial deletion of nucleotides 35,302 to 35,642 of the sequence shown in SEQ ID NO:1,

and wherein the vector comprises at least nucleotides 2 to 36,518 of the sequence set forth in SEQ ID NO:1. The partially deleted E4 can comprise an E4 deletion of at least nucleotides 34,980 to 36,516 of the sequence shown in SEQ ID NO:1, and wherein the vector comprises at least nucleotides 2 to 36,518 of the sequence set forth in SEQ ID NO:1. The partially deleted E4 can comprise an E4 deletion of at least nucleotides 34,979 to 35,642 of the sequence shown in SEQ ID NO:1, and wherein the vector comprises at least nucleotides 2 to 36,518 of the sequence set forth in SEQ ID NO:1. The partially deleted E4 can comprise an E4 deletion of at least a partial deletion of E4Orf2, a fully deleted E4Orf3, and at least a partial deletion of E4Orf4. The partially deleted E4 can comprise an E4 deletion of at least a partial deletion of E4Orf2, at least a partial deletion of E4Orf3, and at least a partial deletion of E4Orf4. The partially deleted E4 can comprise an E4 deletion of at least a partial deletion of E4Orf1, a fully deleted E4Orf2, and at least a partial deletion of E4Orf3. The partially deleted E4 can comprise an E4 deletion of at least a partial deletion of E4Orf2 and at least a partial deletion of E4Orf3. The partially deleted E4 can comprise an E4 deletion between the start site of E4Orf1 to the start site of E4Orf5. The partially deleted E4 can be an E4 deletion adjacent to the start site of E4Orf1. The partially deleted E4 can be an E4 deletion adjacent to the start site of E4Orf2. The partially deleted E4 can be an E4 deletion adjacent to the start site of E4Orf3. The partially deleted E4 can be an E4 deletion adjacent to the start site of E4Orf4. The E4 deletion can be at least 50, at least 100, at least 200, at least 300, at least 400, at least 500, at least 600, at least 700, at least 800, at least 900, at least 1000, at least 1100, at least 1200, at least 1300, at least 1400, at least 1500, at least 1600, at least 1700, at least 1800, at least 1900, or at least 2000 nucleotides. The E4 deletion can be at least 700 nucleotides. The E4 deletion can be at least 1500 nucleotides. The E4 deletion can be 50 or less, 100 or less, 200 or less, 300 or less, 400 or less, 500 or less, 600 or less, 700 or less, 800 or less, 900 or less, 1000 or less, 1100 or less, 1200 or less, 1300 or less, 1400 or less, 1500 or less, 1600 or less, 1700 or less, 1800 or less, 1900 or less, or 2000 or less nucleotides. The E4 deletion can be 750 nucleotides or less. The E4 deletion can be at least 1550 nucleotides or less.

[00328] The partially deleted E4 gene can be the E4 gene sequence shown in SEQ ID NO:1 that lacks at least nucleotides 34,916 to 35,642 of the sequence shown in SEQ ID NO:1. The partially deleted E4 gene can be the E4 gene sequence shown in SEQ ID NO:1 that lacks the E4 gene sequence shown in SEQ ID NO:1 and that lacks at least nucleotides 34,916 to 34,942, nucleotides 34,952 to 35,305 of the sequence shown in SEQ ID NO:1, and nucleotides 35,302 to 35,642 of the sequence shown in SEQ ID NO:1. The partially deleted E4 gene can be the E4 gene sequence shown in SEQ ID NO:1 and that lacks at least nucleotides 34,980 to 36,516 of the sequence shown in SEQ ID NO:1. The partially deleted E4 gene can be the E4 gene sequence shown in SEQ ID NO:1 and that lacks at least nucleotides 34,979 to 35,642 of the sequence

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

[00329] The partially deleted E4 gene can be the E4 gene sequence shown in SEQ ID NO:1 that lacks at least nucleotides 34,916 to 35,642 of the sequence shown in SEQ ID NO:1, nucleotides 2 to 34,916 of the sequence shown in SEQ ID NO:1, wherein the partially deleted E4 gene is 3' of the nucleotides 2 to 34,916, the nucleotides 2 to 34,916 additionally lack nucleotides 577 to 3403 of the sequence shown in SEQ ID NO:1 corresponding to an E1 deletion and lack nucleotides 27,125 to 31,825 of the sequence shown in SEQ ID NO:1 corresponding to an E3 deletion, and have nucleotides 35,643 to 36,518 of the sequence shown in SEQ ID NO:1, and wherein the partially deleted E4 gene is 5' of the nucleotides 35,643 to 36,518.

[00330] Also disclosed herein is a host cell transfected with a vector disclosed herein such as a C68 vector engineered to expression an antigen cassette. Also disclosed herein is a human cell that expresses a selected gene introduced therein through introduction of a vector disclosed herein into the cell.

[00331] Also disclosed herein is a method for delivering an antigen cassette to a mammalian cell comprising introducing into said cell an effective amount of a vector disclosed herein such as a C68 vector engineered to expression the antigen cassette.

[00332] Also disclosed herein is a method for producing an antigen comprising introducing a vector disclosed herein into a mammalian cell, culturing the cell under suitable conditions and producing the antigen.

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

[00333] To generate recombinant chimpanzee adenoviruses (Ad) deleted in any of the genes described herein, the function of the deleted gene region, if essential to the replication and infectivity of the virus, can be supplied to the recombinant virus by a helper virus or cell line, i.e., a complementation or packaging cell line. For example, to generate a replication-defective chimpanzee adenovirus vector, a cell line can be used which expresses the E1 gene products of the human or chimpanzee adenovirus; such a cell line can include HEK293 or variants thereof. The protocol for the generation of the cell lines expressing the chimpanzee E1 gene products (Examples 3 and 4 of USPN 6,083,716) can be followed to generate a cell line which expresses any selected chimpanzee adenovirus gene.

[00334] An AAV augmentation assay can be used to identify a chimpanzee adenovirus E1-expressing cell line. This assay is useful to identify E1 function in cell lines made by using the E1 genes of other uncharacterized adenoviruses, e.g., from other species. That assay is described in Example 4B of USPN 6,083,716.

[00335] A selected chimpanzee adenovirus gene, e.g., E1, can be under the transcriptional control of a promoter for expression in a selected parent cell line. Inducible or constitutive promoters can be employed for this purpose. Among inducible promoters are included the sheep metallothionine promoter, inducible by zinc, or the mouse mammary tumor virus (MMTV) promoter, inducible by a glucocorticoid, particularly, dexamethasone. Other inducible promoters, such as those identified in International patent application WO95/13392, incorporated by reference herein can also be used in the production of packaging cell lines. Constitutive promoters in control of the expression of the chimpanzee adenovirus gene can be employed also.

[00336] A parent cell can be selected for the generation of a novel cell line expressing any desired C68 gene. Without limitation, such a parent cell line can be HeLa [ATCC Accession No.

CCL 2], A549 [ATCC Accession No. CCL 185], KB [CCL 17], Detroit [e.g., Detroit 510, CCL 72] and WI-38 [CCL 75] cells. Other suitable parent cell lines can be obtained from other sources. Parent cell lines can include CHO, HEK293 or variants thereof, 911, HeLa, A549, LP-293, PER.C6, or AE1-2a.

[00337] An E1-expressing cell line can be useful in the generation of recombinant chimpanzee adenovirus E1 deleted vectors. Cell lines constructed using essentially the same procedures that express one or more other chimpanzee adenoviral gene products are useful in the generation of recombinant chimpanzee adenovirus vectors deleted in the genes that encode those products. Further, cell lines which express other human Ad E1 gene products are also useful in generating chimpanzee recombinant Ads.

V.E.3. Recombinant Viral Particles as Vectors

[00338] The compositions disclosed herein can comprise viral vectors, that deliver at least one antigen to cells. Such vectors comprise a chimpanzee adenovirus DNA sequence such as C68 and an antigen cassette operatively linked to regulatory sequences which direct expression of the cassette. The C68 vector is capable of expressing the cassette in an infected mammalian cell. The C68 vector can be functionally deleted in one or more viral genes. An antigen cassette comprises at least one antigen under the control of one or more regulatory sequences such as a promoter. Optional helper viruses and/or packaging cell lines can supply to the chimpanzee viral vector any necessary products of deleted adenoviral genes.

[00339] The term "functionally deleted" means that a sufficient amount of the gene region is removed or otherwise altered, e.g., by mutation or modification, so that the gene region is no longer capable of producing one or more functional products of gene expression. Mutations or modifications that can result in functional deletions include, but are not limited to, nonsense mutations such as introduction of premature stop codons and removal of canonical and non-canonical start codons, mutations that alter mRNA splicing or other transcriptional processing, or combinations thereof. If desired, the entire gene region can be removed.

[00340] Modifications of the nucleic acid sequences forming the vectors disclosed herein, including sequence deletions, insertions, and other mutations may be generated using standard molecular biological techniques and are within the scope of this invention.

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

[00341] The chimpanzee adenovirus C68 vectors useful in this invention include recombinant, defective adenoviruses, that is, chimpanzee adenovirus sequences functionally deleted in the E1a or E1b genes, and optionally bearing other mutations, e.g., temperature-sensitive mutations or deletions in other genes. It is anticipated that these chimpanzee sequences are also useful in

forming hybrid vectors from other adenovirus and/or adeno-associated virus sequences. Homologous adenovirus vectors prepared from human adenoviruses are described in the published literature [see, for example, Kozarsky I and II, cited above, and references cited therein, U.S. Pat. No. 5,240,846].

[00342] In the construction of useful chimpanzee adenovirus C68 vectors for delivery of an antigen cassette to a human (or other mammalian) cell, a range of adenovirus nucleic acid sequences can be employed in the vectors. A vector comprising minimal chimpanzee C68 adenovirus sequences can be used in conjunction with a helper virus to produce an infectious recombinant virus particle. The helper virus provides essential gene products required for viral infectivity and propagation of the minimal chimpanzee adenoviral vector. When only one or more selected deletions of chimpanzee adenovirus genes are made in an otherwise functional viral vector, the deleted gene products can be supplied in the viral vector production process by propagating the virus in a selected packaging cell line that provides the deleted gene functions in trans.

V.E.5. Recombinant Minimal Adenovirus

[00343] A minimal chimpanzee Ad C68 virus is a viral particle containing just the adenovirus cis-elements necessary for replication and virion encapsidation. That is, the vector contains the cis-acting 5' and 3' inverted terminal repeat (ITR) sequences of the adenoviruses (which function as origins of replication) and the native 5' packaging/enhancer domains (that contain sequences necessary for packaging linear Ad genomes and enhancer elements for the E1 promoter). See, for example, the techniques described for preparation of a "minimal" human Ad vector in International Patent Application WO96/13597 and incorporated herein by reference.

V.E.6. Other Defective Adenoviruses

[00344] Recombinant, replication-deficient adenoviruses can also contain more than the minimal chimpanzee adenovirus sequences. These other Ad vectors can be characterized by deletions of various portions of gene regions of the virus, and infectious virus particles formed by the optional use of helper viruses and/or packaging cell lines.

[00345] As one example, suitable vectors may be formed by deleting all or a sufficient portion of the C68 adenoviral immediate early gene E1a and delayed early gene E1b, so as to eliminate their normal biological functions. Replication-defective E1-deleted viruses are capable of replicating and producing infectious virus when grown on a chimpanzee adenovirus-transformed, complementation cell line containing functional adenovirus E1a and E1b genes which provide the corresponding gene products in trans. Based on the homologies to known adenovirus sequences, it is anticipated that, as is true for the human recombinant E1-deleted adenoviruses of

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

[00346] As another example, all or a portion of the C68 adenovirus delayed early gene E3 can be eliminated from the chimpanzee adenovirus sequence which forms a part of the recombinant virus.

[00347] Chimpanzee adenovirus C68 vectors can also be constructed having a deletion of the E4 gene. Still another vector can contain a deletion in the delayed early gene E2a.

[00348] Deletions can also be made in any of the late genes L1 through L5 of the chimpanzee C68 adenovirus genome. Similarly, deletions in the intermediate genes IX and IVa2 can be useful for some purposes. Other deletions may be made in the other structural or non-structural adenovirus genes.

[00349] The above discussed deletions can be used individually, i.e., an adenovirus sequence can contain deletions of E1 only. Alternatively, deletions of entire genes or portions thereof effective to destroy or reduce their biological activity can be used in any combination. For example, in one exemplary vector, the adenovirus C68 sequence can have deletions of the E1 genes and the E4 gene, or of the E1, E2a and E3 genes, or of the E1 and E3 genes, or of E1, E2a and E4 genes, with or without deletion of E3, and so on. As discussed above, such deletions can be used in combination with other mutations, such as temperature-sensitive mutations, to achieve a desired result.

[00350] The cassette comprising antigen(s) be inserted optionally into any deleted region of the chimpanzee C68 Ad virus. Alternatively, the cassette can be inserted into an existing gene region to disrupt the function of that region, if desired.

V.E.7. Helper Viruses

[00351] Depending upon the chimpanzee adenovirus gene content of the viral vectors employed to carry the antigen cassette, a helper adenovirus or non-replicating virus fragment can be used to provide sufficient chimpanzee adenovirus gene sequences to produce an infective recombinant viral particle containing the cassette.

[00352] Useful helper viruses contain selected adenovirus gene sequences not present in the adenovirus vector construct and/or not expressed by the packaging cell line in which the vector is transfected. A helper virus can be replication-defective and contain a variety of adenovirus genes in addition to the sequences described above. The helper virus can be used in combination with the E1-expressing cell lines described herein.

[00353] For C68, the "helper" virus can be a fragment formed by clipping the C terminal end of the C68 genome with SspI, which removes about 1300 bp from the left end of the virus. This clipped virus is then co-transfected into an E1-expressing cell line with the plasmid DNA, thereby forming the recombinant virus by homologous recombination with the C68 sequences in the plasmid.

[00354] Helper viruses can also be formed into poly-cation conjugates as described in Wu et al, J. Biol. Chem., 264:16985-16987 (1989); K. J. Fisher and J. M. Wilson, Biochem. J., 299:49 (Apr. 1, 1994). Helper virus can optionally contain a reporter gene. A number of such reporter genes are known to the art. The presence of a reporter gene on the helper virus which is different from the antigen cassette on the adenovirus vector allows both the Ad vector and the helper virus to be independently monitored. This second reporter is used to enable separation between the resulting recombinant virus and the helper virus upon purification.

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

[00355] Assembly of the selected DNA sequences of the adenovirus, the antigen cassette, and other vector elements into various intermediate plasmids and shuttle vectors, and the use of the plasmids and vectors to produce a recombinant viral particle can all be achieved using conventional techniques. Such techniques include conventional cloning techniques of cDNA, in vitro recombination techniques (e.g., Gibson assembly), use of overlapping oligonucleotide sequences of the adenovirus genomes, polymerase chain reaction, and any suitable method which provides the desired nucleotide sequence. Standard transfection and co-transfection techniques are employed, e.g., CaPO₄ precipitation techniques or liposome-mediated transfection methods such as lipofectamine. Other conventional methods employed include homologous recombination of the viral genomes, plaquing of viruses in agar overlay, methods of measuring signal generation, and the like.

[00356] For example, following the construction and assembly of the desired antigen cassette-containing viral vector, the vector can be transfected in vitro in the presence of a helper virus into the packaging cell line. Homologous recombination occurs between the helper and the vector sequences, which permits the adenovirus-antigen sequences in the vector to be replicated and packaged into virion capsids, resulting in the recombinant viral vector particles.

[00357] The resulting recombinant chimpanzee C68 adenoviruses are useful in transferring an antigen cassette to a selected cell. In in vivo experiments with the recombinant virus grown in the packaging cell lines, the E1-deleted recombinant chimpanzee adenovirus demonstrates utility in transferring a cassette to a non-chimpanzee, preferably a human, cell.

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

[00358] The resulting recombinant chimpanzee C68 adenovirus containing the antigen cassette (produced by cooperation of the adenovirus vector and helper virus or adenoviral vector and packaging cell line, as described above) thus provides an efficient gene transfer vehicle which can deliver antigen(s) to a subject in vivo or ex vivo.

[00359] The above-described recombinant vectors are administered to humans according to published methods for gene therapy. A chimpanzee viral vector bearing an antigen cassette can be administered to a patient, preferably suspended in a biologically compatible solution or pharmaceutically acceptable delivery vehicle. A suitable vehicle includes sterile saline. Other aqueous and non-aqueous isotonic sterile injection solutions and aqueous and non-aqueous sterile suspensions known to be pharmaceutically acceptable carriers and well known to those of skill in the art may be employed for this purpose.

[00360] The chimpanzee adenoviral vectors are administered in sufficient amounts to transduce the human cells and to provide sufficient levels of antigen transfer and expression to provide a therapeutic benefit without undue adverse or with medically acceptable physiological effects, which can be determined by those skilled in the medical arts. Conventional and pharmaceutically acceptable routes of administration include, but are not limited to, direct delivery to the liver, intranasal, intravenous, intramuscular, subcutaneous, intradermal, oral and other parental routes of administration. Routes of administration may be combined, if desired.

[00361] Dosages of the viral vector will depend primarily on factors such as the condition being treated, the age, weight and health of the patient, and may thus vary among patients. The dosage will be adjusted to balance the therapeutic benefit against any side effects and such dosages may vary depending upon the therapeutic application for which the recombinant vector is employed. The levels of expression of antigen(s) can be monitored to determine the frequency of dosage administration.

[00362] Recombinant, replication defective adenoviruses can be administered in a "pharmaceutically effective amount", that is, an amount of recombinant adenovirus that is effective in a route of administration to transfect the desired cells and provide sufficient levels of expression of the selected gene to provide a vaccinal benefit, i.e., some measurable level of protective immunity. C68 vectors comprising an antigen cassette can be co-administered with adjuvant. Adjuvant can be separate from the vector (e.g., alum) or encoded within the vector, in particular if the adjuvant is a protein. Adjuvants are well known in the art.

[00363] Conventional and pharmaceutically acceptable routes of administration include, but are not limited to, intranasal, intramuscular, intratracheal, subcutaneous, intradermal, rectal, oral

and other parental routes of administration. Routes of administration may be combined, if desired, or adjusted depending upon the immunogen or the disease. For example, in prophylaxis of rabies, the subcutaneous, intratracheal and intranasal routes are preferred. The route of administration primarily will depend on the nature of the disease being treated.

[00364] The levels of immunity to antigen(s) can be monitored to determine the need, if any, for boosters. Following an assessment of antibody titers in the serum, for example, optional booster immunizations may be desired

VI. Therapeutic and Manufacturing Methods

[00365] 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.

[00366] 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.

[00367] 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.

[00368] In some aspects, a subject has been diagnosed with an infection or is at risk of an infection (*e.g.*, 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).

[00369] 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. An antigen can be administered in an amount sufficient to stimulate both a T cell response and a B cell response

[00370] 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.

[00371] 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.

[00372] 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.

[00373] 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.

[00374] 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.

[00375] 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 particularly antigen or pathway of this antigen can be included.

[00376] 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 subject 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.

[00377] 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).

[00378] The pharmaceutical compositions (*e.g.*, vaccine compositions) for therapeutic treatment are intended for parenteral, topical, nasal, oral or local administration. A pharmaceutical compositions can be administered parenterally, *e.g.*, intravenously, subcutaneously, intradermally, or intramuscularly. The compositions can be administered 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.

[00379] 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.

[00380] 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.

[00381] 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.

[00382] 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).

[00383] 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.

[00384] 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.

[00385] 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.

[00386] 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.

[00387] 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.

[00388] 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.

VII. Antigen Use and Administration

[00389] 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. Vaccination methods, protocols, and schedules that can be used include, but are not limited to, those described in international application publication WO2021092095, herein incorporated by reference for all purposes.

[00390] A priming vaccine, can 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 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).

[00391] 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. 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).

[00392] A priming vaccine can be injected (e.g., intramuscularly) in a subject. Bilateral injections per dose can be used. For example, one or more injections of ChAdV68 (C68) can be used (e.g., total dose 1×10^{12} viral particles); one or more injections of SAM vectors at low vaccine dose selected from the range 0.001 to 1 ug RNA, in particular 0.1 or 1 ug can be used; or one or more injections of SAM vectors at high vaccine dose selected from the range 1 to 100 ug RNA, in particular 10 or 100 ug can be used.

[00393] A vaccine boost (boosting vaccine) can be injected (e.g., intramuscularly) after prime vaccination. A boosting vaccine can be administered about every 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 weeks, e.g., every 4 weeks and/or 8 weeks after the prime. Bilateral injections per dose can be used. For example, one or more injections of ChAdV68 (C68) can be used (e.g., total dose 1×10^{12} viral particles); one or more injections of SAM vectors at low vaccine dose selected from the range 0.001 to 1 ug RNA, in particular 0.1 or 1 ug can be used; or one or more injections of SAM vectors at high vaccine dose selected from the range 1 to 100 ug RNA, in particular 10 or 100 ug can be used.

[00394] 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.

[00395] 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.

[00396] Immune monitoring can be performed before, during, and/or after vaccine administration. Such monitoring can inform safety and efficacy, among other parameters.

[00397] 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).

[00398] 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.

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

[00400] Disease status of a subject can be monitored following administration of any of the vaccine compositions described herein. For example, disease status may be monitored using isolated cell-free DNA (cfDNA) from a subject. In addition, the efficacy of a vaccine therapy may be monitored using isolated cfDNA from a subject. cfDNA monitoring can include the steps of: a. isolating or having isolated cfDNA from a subject; b. sequencing or having sequenced the isolated cfDNA; c. determining or having determined a frequency of one or more mutations in the cfDNA relative to a wild-type germline nucleic acid sequence of the subject, and d. assessing or having assessed from step (c) the status of a disease in the subject. The method can also include, following step (c) above, d. performing more than one iteration of steps (a)-(c) for the given subject and comparing the frequency of the one or more mutations determined in the more than one iterations; and f. assessing or having assessed from step (d) the status of a disease in the

subject. The more than one iterations can be performed at different time points, such as a first iteration of steps (a)-(c) performed prior to administration of the vaccine composition and a second iteration of steps (a)-(c) is performed subsequent to administration of the vaccine composition. Step (c) can include comparing: the frequency of the one or more mutations determined in the more than one iterations, or the frequency of the one or more mutations determined in the first iteration to the frequency of the one or more mutations determined in the second iteration. An increase in the frequency of the one or more mutations determined in subsequent iterations or the second iteration can be assessed as disease progression. A decrease in the frequency of the one or more mutations determined in subsequent iterations or the second iteration can be assessed as a response. In some aspects, the response is a Complete Response (CR) or a Partial Response (PR). A therapy can be administered to a subject following an assessment step, such as where assessment of the frequency of the one or more mutations in the cfDNA indicates the subject has the disease. The cfDNA isolation step can use centrifugation to separate cfDNA from cells or cellular debris. cfDNA can be isolated from whole blood, such as by separating the plasma layer, buffy coat, and red bloods. cfDNA sequencing can use next generation sequencing (NGS), Sanger sequencing, duplex sequencing, whole-exome sequencing, whole-genome sequencing, de novo sequencing, phased sequencing, targeted amplicon sequencing, shotgun sequencing, or combinations thereof, and may include enriching the cfDNA for one or more polynucleotide regions of interest prior to sequencing (*e.g.*, polynucleotides known or suspected to encode the one or more mutations, coding regions, and/or tumor exome polynucleotides). Enriching the cfDNA may include hybridizing one or more polynucleotide probes, which may be modified (*e.g.*, biotinylated), to the one or more polynucleotide regions of interest. In general, any number of mutations may be monitored simultaneously or in parallel.

VIII. Isolation and Detection of HLA Peptides

[00401] Isolation of HLA-peptide molecules was performed using classic immunoprecipitation (IP) methods after lysis and solubilization of the tissue sample (55-58). A clarified lysate was used for HLA specific IP.

[00402] Immunoprecipitation was performed using antibodies coupled to beads where the antibody is specific for HLA molecules. For a pan-Class I HLA immunoprecipitation, a pan-Class I CR antibody is used, for Class II HLA – DR, an HLA-DR antibody is used. Antibody is covalently attached to NHS-sepharose beads during overnight incubation. After covalent attachment, the beads were washed and aliquoted for IP. (59, 60) Immunoprecipitations can also be performed with antibodies that are not covalently attached to beads. Typically this is done using sepharose or magnetic beads coated with Protein A and/or Protein G to hold the antibody

to the column. Some antibodies that can be used to selectively enrich MHC/peptide complex are listed below.

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

[00403] The clarified tissue lysate is added to the antibody beads for the immunoprecipitation. After immunoprecipitation, the beads are removed from the lysate and the lysate stored for additional experiments, including additional IPs. The IP beads are washed to remove non-specific binding and the HLA/peptide complex is eluted from the beads using standard techniques. The protein components are removed from the peptides using a molecular weight spin column or C18 fractionation. The resultant peptides are taken to dryness by SpeedVac evaporation and in some instances are stored at -20C prior to MS analysis.

[00404] Dried peptides are reconstituted in an HPLC buffer suitable for reverse phase chromatography and loaded onto a C-18 microcapillary HPLC column for gradient elution in a Fusion Lumos mass spectrometer (Thermo). MS1 spectra of peptide mass/charge (m/z) were collected in the Orbitrap detector at high resolution followed by MS2 low resolution scans collected in the ion trap detector after HCD fragmentation of the selected ion. Additionally, MS2 spectra can be obtained using either CID or ETD fragmentation methods or any combination of the three techniques to attain greater amino acid coverage of the peptide. MS2 spectra can also be measured with high resolution mass accuracy in the Orbitrap detector with targeted method known as parallel reaction monitoring. In targeted PRM, specific peptide precursor ions are isolated in the Orbitrap detector and all resulting HCD fragmentation ions are scanned across the elution of the peptide peak. This enables both peptide identification and quantitation of endogenous peptide in the presence of a co-injected stable isotopically labeled peptide standard..

[00405] MS2 spectra from each analysis are searched against a protein database using Comet (61, 62) and the peptide identification are scored using Percolator (63-65). Additional sequencing is performed using PEAKS studio (Bioinformatics Solutions Inc.) and other search engines or sequencing methods can be used including spectral matching and *de novo* sequencing (97). Targeted MS1 and MS2 spectra are processed through Skyline (104).

VIII.B.1. MS limit of detection studies in support of comprehensive HLA peptide sequencing

[00406] Using the peptide YVYVADVAAK it was determined what the limits of detection are using different amounts of peptide loaded onto the LC column. The amounts of peptide

tested were 1 pmol, 100fmol, 10 fmol, 1 fmol, and 100amol. (Table 1) These results indicate that the lowest limit of detection (LoD) is in the attomol range (10^{-18}), that the dynamic range spans five orders of magnitude, and that the signal to noise appears sufficient for sequencing at low femtomol ranges (10^{-15}).

Table 1

Peptide m/z	Loaded on Column	Copies/Cell in 1e9cells
566.830	1 pmol	600
562.823	100 fmol	60
559.816	10 fmol	6
556.810	1 fmol	0.6
553.802	100 amol	0.06

IX. Presentation Model

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

X. Training Module

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

XI. Prediction Module

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

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

XI.B.Cassette Design Module

XI.B.1 Overview

[00411] A cassette design module can be used to generate a vaccine cassette sequence based on selected candidate peptides for injection into a patient. Various cassette design modules are known to those skilled in the art, for example the cassette design modules described in more detail in US Pat No. 10,055,540, US Application Pub. No. US20200010849A1, and international patent application publications WO/2018/195357 and WO/2018/208856, each herein incorporated by reference, in their entirety, for all purposes.

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

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

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

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

Such reactions are often times undesirable because T-cells reactive to the junction epitopes have no therapeutic benefit, and may diminish the immune response to the selected therapeutic epitopes in the cassette by antigenic competition.⁷⁶

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

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

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

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

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

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

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

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

[00422] An illustrative non-limiting cassette of concatenated KRAS-associated MHC class I neoepitopes that are linked through their native flanking sequences, includes 4 iterations for each of the KRAS neoepitopes having the mutations KRAS G12C, KRAS G12D, KRAS G12V, and KRAS Q61H, and has been ordered to minimize potential junctional epitopes is represented by the amino acid sequence shown in SEQ ID NO: 65 and having the order of KRAS-associated neoepitopes: G12C G12D Q61H G12D G12V G12C Q61H G12D G12V G12C Q61H G12D G12V Q61H G12V G12C.

[00423] Shared (neo)antigen sequences for inclusion in a shared antigen vaccine and appropriate patients for treatment with such vaccine can be chosen by one of skill in the art, e.g., as described in US App. No. 17/058,128, herein incorporated by reference for all purposes. Mass spectrometry (MS) validation of candidate shared (neo)antigens can be performed as part of the selection process.

XIII. Example Computer

[00424] A computer can be used for any of the computational methods described herein. One skilled in the art will recognize a computer can have different architectures. Examples of computers are known to those skilled in the art, for example the computers described in more detail in US Pat No. 10,055,540, US Application Pub. No. US20200010849A1, and international patent application publications WO/2018/195357 and WO/2018/208856, each herein incorporated by reference, in their entirety, for all purposes.

XIV. Examples

[00425] Below are examples of specific embodiments for carrying out the present invention. The examples are offered for illustrative purposes only, and are not intended to limit the scope of the present invention in any way. Efforts have been made to ensure accuracy with respect to numbers used (e.g., amounts, temperatures, etc.), but some experimental error and deviation should, of course, be allowed for.

[00426] The practice of the present invention will employ, unless otherwise indicated, conventional methods of protein chemistry, biochemistry, recombinant DNA techniques and pharmacology, within the skill of the art. Such techniques are explained fully in the literature. See, e.g., T.E. Creighton, *Proteins: Structures and Molecular Properties* (W.H. Freeman and Company, 1993); A.L. Lehninger, *Biochemistry* (Worth Publishers, Inc., current addition);

Sambrook, et al., *Molecular Cloning: A Laboratory Manual* (2nd Edition, 1989); *Methods In Enzymology* (S. Colowick and N. Kaplan eds., Academic Press, Inc.); *Remington's Pharmaceutical Sciences*, 18th Edition (Easton, Pennsylvania: Mack Publishing Company, 1990); Carey and Sundberg *Advanced Organic Chemistry 3rd Ed.* (Plenum Press) Vols A and B(1992).

XIV.A. Evaluation of Cassettes with Repeated Neoepitopes and/or Removal of Immunodominant Neoepitopes Overview

[00427] Through vaccination, multiple class I MHC restricted neoantigens that stimulate the corresponding cellular immune response(s) can be delivered. Several vaccine cassettes were engineered to encode multiple neoepitopes, specifically multiple distinct KRAS-associated neoepitopes, as a single gene product where the epitopes were embedded within their natural, surrounding peptide sequence. Various cassettes designs feature multiple copies (*i.e.*, iterations) of one or more neoepitopes. Various cassettes designs also feature removing immunodominant neoepitopes.

XIV.B. Repeated Neoepitopes and/or Removal of Immunodominant Neoepitopes Evaluation Materials and Methods

Antigens

[00428] The epitope-containing 25mer amino acid sequences (*i.e.*, epitope flanked by native N and C terminal amino acid linkers) used in the examples below are presented in **Table 2A** were used as described below. The antigen-encoding sequences of the cassettes for the various constructs were constructed by directly linking each 25mer sequence to each other (*i.e.*, no additional amino acids in between consecutive 25mer sequences) in the order and number described in the examples below, *e.g.*, see **FIGs. 1A, 2A, 3A, and 4A**. The cassettes containing the full-length antigen-encoding sequences containing the multiple distinct epitopes linked together, as well as universal MHC class II antigens tetanus toxoid and PADRE (bolded sequence), are presented in **Table 2B**. The complete exogenous nucleotide insertions into the vectors include from 5' to 3': a Kozak sequence (GCCACC), nucleotides encoding three amino acids MAG (ATGGCCGGG), one of the cassette sequences of **Table 2B**, and two stop codons (TAATGA). Further cassettes were generated with 4 iterations of individual KRAS mutations G12V, G12C, and G12D (*i.e.*, separate concatemers of SEQ ID NOs 57, 58, or 59).

Table 2A – 25mer Neoepitope Containing Sequences

Antigen Name	25mer Sequence	SEQ ID NO:
KRAS G12C	MTEYKLVVVGACGVGKSALTIQLIQ	57
KRAS G12D	MTEYKLVVVGADGVGKSALTIQLIQ	58
KRAS G12V	MTEYKLVVVGAVGVGKSALTIQLIQ	59
KRAS Q61H	ETCLLDILDITAGHEEYSAMRDQYMR	60
TP53 R213L	LRVEYLDDRNTFLHSVVVPYEPPEV	61
TP53 S127Y	HSGTAKSVTCTYYPALNKMFCQLAK	62

Table 2B – Full-Length Multi-Neoepitope Sequences in Cassettes

1X (20x1) Cassette (SEQ ID NO: 63)
MTEYKLVVVGAGDVGKSALTIQLIQETCLLDILDITAGKEEYSAMRDQYMRMCNSSCM GGMNRMPIITITLEDSSSGIHSGATTTAPPLSGKGNPEEEDVSGIHSGATTTAPFLSGKG NPEEEDVEILDEAYVMAYVMAGVGSPYVSRLLMTEYKLVVVGADGVGKSALTIQLIQ ETCLLDILDITAGREEYSAMRDQYMRSYLDSGIHSGATATAPSLSGKGNPEKSVTCTYSP ALNNMFCQLAKTCPVQMTEYKLVVVGAAGVGKSALTIQLIQETCLLDILDITAGLEEYS AMRDQYMRRLRVEYLDDRNTFLHSVVVPYEPPEVDGQITVGQRIGSVSFGTVYKKGKWH GMTEYKLVVVGAVGVGKSALTIQLIQETCLLDILDITAGHEEYSAMRDQYMRWQQQSY LDSGIHFGATTTAPSLSGKHSGTAKSVTCTYYPALNKMFCQLAKKSVTCTYSPALNEM FCQLAKTCPVQMTEYKLVVVGACGVGKSALTIQLIQGPGPGAKFVAAWTLKAAAGP GPGQYIKANSKFIGITELGPGPG
KRAS 2X (8x2) Cassette (SEQ ID NO: 64)
MTEYKLVVVGACGVGKSALTIQLIQETCLLDILDITAGHEEYSAMRDQYMRMTEYKL VVVGAGDVGKSALTIQLIQETCLLDILDITAGREEYSAMRDQYMRMTEYKLVVVGADGV GKSALTIQLIQETCLLDILDITAGLEEYSAMRDQYMRMTEYKLVVVGAVGVGKSALTIQ LIQETCLLDILDITAGKEEYSAMRDQYMRMTEYKLVVVGACGVGKSALTIQLIQETCLL DILDITAGKEEYSAMRDQYMRMTEYKLVVVGAGDVGKSALTIQLIQETCLLDILDITAGH EEYSAMRDQYMRMTEYKLVVVGAVGVGKSALTIQLIQETCLLDILDITAGREEYSAMR DQYMRMTEYKLVVVGADGVGKSALTIQLIQETCLLDILDITAGLEEYSAMRDQYMRGP GPGAKFVAAWTLKAAAGPGPQYIKANSKFIGITELGPGPG
KRAS 4X (4x4) Cassette (SEQ ID NO: 65)
MTEYKLVVVGACGVGKSALTIQLIQMTEYKLVVVGADGVGKSALTIQLIQETCLLDIL DTAGHEEYSAMRDQYMRMTEYKLVVVGADGVGKSALTIQLIQMTEYKLVVVGAVGV GKSALTIQLIQMTEYKLVVVGACGVGKSALTIQLIQETCLLDILDITAGHEEYSAMRDQY MRMTEYKLVVVGADGVGKSALTIQLIQMTEYKLVVVGAVGVGKSALTIQLIQMTEYK LVVVGACGVGKSALTIQLIQETCLLDILDITAGHEEYSAMRDQYMRMTEYKLVVVGAD GVGKSALTIQLIQMTEYKLVVVGAVGVGKSALTIQLIQETCLLDILDITAGHEEYSAMRD QYMRMTEYKLVVVGAVGVGKSALTIQLIQMTEYKLVVVGACGVGKSALTIQLIQGPG PGAKFVAAWTLKAAAGPGPQYIKANSKFIGITELGPGPG
KRAS 4x1 Cassette (SEQ ID NO: 66)
MTEYKLVVVGACGVGKSALTIQLIQMTEYKLVVVGADGVGKSALTIQLIQETCLLDIL DTAGHEEYSAMRDQYMRMTEYKLVVVGAVGVGKSALTIQLIQGPGPGAKFVAAWT LKAAAGPGPQYIKANSKFIGITELGPGPG
KRAS 4x1 + R213L Cassette (SEQ ID NO: 67)
MTEYKLVVVGACGVGKSALTIQLIQMTEYKLVVVGADGVGKSALTIQLIQETCLLDIL DTAGHEEYSAMRDQYMRMTEYKLVVVGAVGVGKSALTIQLIQLRVEYLDDRNTFLHS VVVPYEPPEVGPGPGAKFVAAWTLKAAAGPGPQYIKANSKFIGITELGPGPG

KRAS 4x1 + S127Y Cassette (SEQ ID NO: 68)
MTEYKLVVVGACGVGKSALTIQLIQMTEYKLVVVGADGVGKSALTIQLIQETCLLDIL DTAGHEEYSAMRDQYMRMTEYKLVVVGAVGVGKSALTIQLIQHSGTAKSVTCTYYPA LNKMFCQLAKGPGPGAKFVAAWTLKAAAGPQYIKANSKFIGITELGPGPG

Adenoviral Vectors

[00429] A modified ChAdV68 vector for the antigen expression system was generated based on AC_000011.1 with E1 (nt 577 to 3403) and E3 (nt 27,125- 31,825) sequences deleted and the corresponding ATCC VR-594 (Independently sequenced Full-Length VR-594 C68 SEQ ID NO:10) nucleotides substituted at five positions. The full-length ChAdVC68 AC_000011.1 sequence with corresponding ATCC VR-594 nucleotides substituted at five positions is referred to as “ChAdV68.5WTnt” (SEQ ID NO:1). Antigen cassettes under the control of the CMV promoter/enhancer were inserted in place of deleted E1 sequences. A representative ChAdV68 vector containing 20 model antigens in an antigen cassette is “ChAdV68.5WTnt.MAG25mer” (SEQ ID NO:2). The vectors featuring antigen cassettes described below having the MAG25mer cassette can be replaced by the antigen cassettes described above, e.g., in Table 2B.

Adenoviral Production in 293F cells

[00430] ChAdV68 virus production was performed in 293F cells grown in 293 FreeStyle™ (ThermoFisher) media in an incubator at 8% CO₂. On the day of infection cells were diluted to 10⁶ cells per mL, with 98% viability and 400 mL were used per production run in 1L Shake flasks (Corning). 4 mL of the tertiary viral stock with a target MOI of >3.3 was used per infection. The cells were incubated for 48-72h until the viability was <70% as measured by Trypan blue. The infected cells were then harvested by centrifugation, full speed bench top centrifuge and washed in 1XPBS, re-centrifuged and then re-suspended in 20 mL of 10mM Tris pH7.4. The cell pellet was lysed by freeze thawing 3X and clarified by centrifugation at 4,300Xg for 5 minutes.

Adenoviral Purification by CsCl centrifugation

[00431] Viral DNA was purified by CsCl centrifugation. Two discontinuous gradient runs were performed. The first to purify virus from cellular components and the second to further refine separation from cellular components and separate defective from infectious particles.

[00432] 10 mL of 1.2 (26.8g CsCl dissolved in 92 mL of 10 mM Tris pH 8.0) CsCl was added to polyallomer tubes. Then 8 mL of 1.4 CsCl (53g CsCl dissolved in 87 mL of 10 mM Tris pH 8.0) was carefully added using a pipette delivering to the bottom of the tube. The clarified virus was carefully layered on top of the 1.2 layer. If needed more 10 mM Tris was added to balance the tubes. The tubes were then placed in a SW-32Ti rotor and centrifuged for 2h 30 min at 10⁰C. The tube was then removed to a laminar flow cabinet and the virus band pulled using an 18

gauge needle and a 10 mL syringe. Care was taken not to remove contaminating host cell DNA and protein. The band was then diluted at least 2X with 10 mM Tris pH 8.0 and layered as before on a discontinuous gradient as described above. The run was performed as described before except that this time the run was performed overnight. The next day the band was pulled with care to avoid pulling any of the defective particle band. The virus was then dialyzed using a Slide-a-Lyzer™ Cassette (Pierce) against ARM buffer (20 mM Tris pH 8.0, 25 mM NaCl, 2.5% Glycerol). This was performed 3X, 1h per buffer exchange. The virus was then aliquoted for storage at -80°C.

Adenoviral Viral Assays

[00433] VP concentration was performed by using an OD 260 assay based on the extinction coefficient of 1.1×10^{12} viral particles (VP) is equivalent to an Absorbance value of 1 at OD260 nm. Two dilutions (1:5 and 1:10) of adenovirus were made in a viral lysis buffer (0.1% SDS, 10 mM Tris pH 7.4, 1mM EDTA). OD was measured in duplicate at both dilutions and the VP concentration/ mL was measured by multiplying the OD260 value X dilution factor X 1.1×10^{12} VP.

[00434] An infectious unit (IU) titer was calculated by a limiting dilution assay of the viral stock. The virus was initially diluted 100X in DMEM/5% NS/ 1X PS and then subsequently diluted using 10-fold dilutions down to 1×10^{-7} . 100 μ L of these dilutions were then added to 293A cells that were seeded at least an hour before at 3×10^5 cells/ well of a 24 well plate. This was performed in duplicate. Plates were incubated for 48h in a CO₂ (5%) incubator at 37 °C. The cells were then washed with 1XPBS and were then fixed with 100% cold methanol (-20 °C). The plates were then incubated at -20 °C for a minimum of 20 minutes. The wells were washed with 1XPBS then blocked in 1XPBS/0.1% BSA for 1 h at room temperature. A rabbit anti-Ad antibody (Abcam, Cambridge, MA) was added at 1:8,000 dilution in blocking buffer (0.25 ml per well) and incubated for 1 h at room temperature. The wells were washed 4X with 0.5 mL PBS per well. A HRP conjugated Goat anti-Rabbit antibody (Bethyl Labs, Montgomery Texas) diluted 1000X was added per well and incubated for 1h prior to a final round of washing. 5 PBS washes were performed and the plates were developed using DAB (Diaminobenzidine tetrahydrochloride) substrate in Tris buffered saline (0.67 mg/mL DAB in 50 mM Tris pH 7.5, 150 mM NaCl) with 0.01% H₂O₂. Wells were developed for 5 min prior to counting. Cells were counted under a 10X objective using a dilution that gave between 4-40 stained cells per field of view. The field of view that was used was a 0.32 mm² grid of which there are equivalent to 625 per field of view on a 24 well plate. The number of infectious viruses/ mL can be determined by the number of stained cells per grid multiplied by the number of grids per field of view multiplied by a dilution factor 10. Similarly, when working with GFP expressing cells fluorescent

can be used rather than capsid staining to determine the number of GFP expressing virions per mL.

SAM Vectors

[00435] A RNA alphavirus backbone for the antigen expression system was generated from a self-replicating Venezuelan Equine Encephalitis (VEE) virus (Kinney, 1986, Virology 152: 400-413) by deleting the structural proteins of VEE located 3' of the 26S sub-genomic promoter (VEE sequences 7544 to 11,175 deleted; numbering based on Kinney *et al* 1986; SEQ ID NO:6). To generate the self-amplifying mRNA ("SAM") vector, the deleted sequences are replaced by antigen sequences. A representative SAM vector containing 20 model antigens is "VEE-MAG25mer" (SEQ ID NO:4). The vectors featuring the antigen cassettes described having the MAG25mer cassette can be replaced by the SARS-CoV-2 cassettes and/or full-length proteins described herein.

In vitro transcription to generate SAM

[00436] *For in vivo studies:* SAM vectors were generated as "AU-SAM" vectors. 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 (*e.g.*, the sequence set forth in SEQ ID NO:6; SAM with 7544 to 11,175 deleted without an inserted antigen cassette). Reaction conditions are described below:

- 1x transcription buffer (40 mM Tris-HCL [pH7.9], 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 (E2040S); 0.025 mg/mL DNA transcription template (linearized by restriction digest); 8 mM CleanCap Reagent AU (Cat. No. N-7114) and 10 mM each of ATP, cytidine triphosphate (CTP), GTP, and uridine triphosphate (UTP)
- Transcription reactions were incubated at 37°C for 2 hr and treated with final 2 U DNase I (AM2239) /0.001 mg DNA transcription template in DNase I buffer for 1 hr at 37°C.
- SAM was purified by RNeasy Maxi (QIAGEN, 75162)

[00437] Alternatively to co-transcriptional addition of a 5' cap structure, a 7-methylguanosine or a related 5' cap structure can be enzymatically added following transcription using a vaccinia capping system (NEB Cat. No. M2080) containing mRNA 2'-O-methyltransferase and S-Adenosyl methionine.

Immunizations

[00438] For ChAdV68 vaccines, transgenic mice expressing a chimeric HLA-A11:01 (Taconic Model #9660 – [CB6F1-Tg(HLA-A*1101/H2-Kb)A11.01]) were injected with 8×10^{10}

viral particles (VP) or 5×10^{10} VP, as indicated, in 100 μ L volume, bilateral intramuscular injection (50 μ L per leg).

[00439] For SAM vaccines, 10 μ g of RNA-LNP complexes in 100 μ L volume were administered as a bilateral intramuscular injection (50 μ L per leg).

Splenocyte dissociation

[00440] Splenocytes were isolated 14 days 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 NH_4Cl , 10mM KHCO_3 , 0.1mM Na_2EDTA). 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

[00441] ELISPOT analysis was performed according to ELISPOT harmonization guidelines {DOI: 10.1038/nprot.2015.068} with the mouse IFN γ ELISpotPLUS kit (MABTECH). 1×10^5 splenocytes were incubated with 10 μ M of the indicated 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.

XIV.C. Repeated Neopitope Evaluation Results

[00442] Vaccine efficacy with cassettes having repeated (*i.e.*, iterated) neopeptides was compared against cassettes having only a single copy of a given neopeptide. Mice engineered to express human HLA-A11:01 were immunized with 8×10^{10} VP using the ChAdV68 delivery vectors described below and efficacy was assessed by IFN γ ELISpot.

[00443] As illustrated in FIG. 1A, a series of ChAdV68 delivery vectors were designed to assess efficacy of cassettes featuring either a single copy of KRAS neopeptides G12C, G12V,

G12D, and Q61H (“KRAS 1X (20x1)”); cassette = SEQ ID NO: 63), 2 repeats of the KRAS G12C, G12V, G12D, and Q61H neoepitopes and 2 repeats of additional KRAS neoepitopes (“KRAS 2X (8x2)”); cassette = SEQ ID NO: 64), or 4 repeats of the KRAS neoepitopes (“KRAS 4X (4x4)”); cassette = SEQ ID NO: 65). HLA-A11:01 was previously predicted to present KRAS G12C, G12V, and G12D neoepitopes (data not shown).

[00444] As shown in **FIGs. 1B** and presented in **Table 3**, vaccination with vectors featuring multiple repeats of the neoepitopes demonstrated increased spot forming colonies (SFC), indicating repetition of epitope-encoding sequences in a cassette led to an increased antigen-specific immune response against KRAS neoepitope G12C. However, upon further observation that there was a response to the G12C peptide with the G12V 1x4 cassette (only expressing G12V) and none with the G12C 1x4 cassette (only expressing G12C), the observed response to the G12C peptide with the 4x4 cassette was likely driven by the G12V epitope. Mass spec analysis confirmed that the G12C peptide was contaminated with G12V peptide.

[00445] As shown in **FIG. 1C** and **FIG. 1D** and presented in **Table 3A**, vaccination with vectors featuring multiple repeats of the neoepitopes demonstrated increased spot forming colonies (SFC), indicating repetition of epitope-encoding sequences in a cassette led to an increased antigen-specific immune response against KRAS neoepitopes, specifically G12V and G12D neoepitopes, using a ChAdV68 delivery system.

Table 3 – ELISpot data for KRAS Neoepitopes G12C, G12V, and G12D +/- Repetition

Antigen Cassette	G12C		G12V		G12D	
	Mean	+/- SE	Mean	+/- SE	Mean	+/- SE
20x1	17	7	3	2	5	3
4x4	2277 ⁺	580	6169 [*]	31	970	864
8x2	66	31	1563	293	97	45

* 5/6 samples were too numerous to count (max value 6200)

+ likely due to contamination with G12V

XIV.D. Immunodominant Epitope Removal Evaluation Results

[00446] Vaccine efficacy with cassettes having a potentially immunodominant TP53 neoepitope capable of stimulating an immune response was compared against cassettes without a TP53 neoepitope capable of stimulating an immune response. Mice engineered to express human HLA-A11:01 were immunized with 5×10^{10} VP using the ChAdV68 delivery vectors described below and efficacy was assessed by ELISpot. HLA-A11:01 was previously predicted to present KRAS G12C, G12V, G12D, and TP53 S127Y neoepitopes (data not shown).

[00447] As illustrated in **FIG. 2A**, a series of ChAdV68 delivery vectors were designed to assess immunodominance of a TP53 epitope. Vector containing only KRAS neoepitopes G12C, G12V, G12D, and Q61H (“KRAS 4x1”; cassette = SEQ ID NO: 66), KRAS neoepitopes in combination with a TP53 R213L neoepitope considered unable to stimulate an immune response in the HLA-A11:01 mouse model as it is predicted to be presented by human HLA-A02:01 and not HLA-A11:01 (“KRAS 4x1 + R213L”; cassette = SEQ ID NO: 67), or KRAS neoepitopes in combination with a TP53 S127Y neoepitope predicted to be presented by HLA-A11:01 (“KRAS 4x1 + S127Y”; cassette = SEQ ID NO: 68).

[00448] As shown in **FIGs. 2B** and presented in **Table 4**, vaccination with vectors including the immunogenic TP53 S127Y epitope demonstrated reduced spot forming colonies (SFC) for KRAS epitope G12C relative to vectors not including a TP53 neoepitope (“4x1”) or including a non-immunogenic TP53 neoepitope (“4x1+R213L”), indicating presence of the TP53 S127Y neoepitope in the vaccine cassette acts as an immunodominant neoepitope decreasing KRAS-specific T cell responses. However, upon further observation that there was a response to the G12C peptide with the G12V 1x4 cassette (only expressing G12V) and none with the G12C 1x4 cassette (only expressing G12C), the observed response to the G12C peptide with the 4x4 cassette was likely driven by the G12V epitope. Mass spec analysis confirmed that the G12C peptide was contaminated with G12V peptide.

[00449] As shown in **FIGs. 2C-E** and presented in **Table 4**, vaccination with vectors including the immunogenic TP53 S127Y epitope demonstrated reduced spot forming colonies (SFC) for KRAS epitopes G12V and G12D relative to vectors not including a TP53 neoepitope (“4x1”) or including a non-immunogenic TP53 neoepitope (“4x1+R213L”), indicating presence of the TP53 S127Y neoepitope in the vaccine cassette acts as an immunodominant neoepitope decreasing KRAS-specific T cell responses.

Table 4 – ELISpot data for KRAS and TP53 Neoepitopes

Antigen Cassette	G12C		G12D		G12V		R213L		S127Y	
	Mean	+/-SE	Mean	+/-SE	Mean	+/-SE	Mean	+/-SE	Mean	+/-SE
4x1	92	41	32	21	1570	833	ND	ND	ND	ND
4x1+R213L	86	74	62	46	773	378	5	3	ND	ND
4x1+S127Y	2	1	7	5	0	0	ND	ND	6118*	264

ND = Not Determined; *3/6 samples were too numerous to count (max value 6200)

XIV.E. Evaluation of Presentation of KRAS Neopeptides by Mass Spectrometry

[00450] Mass spectrometry (MS) validation of candidate KRAS epitopes was performed using targeted mass spectrometry methods. Either donor tumor resections (multi-allelic; allele assignment from EDGE prediction) or single allelic K562 lines engineered to express the indicated cassettes were assessed. Mass spectrometry analysis methods are described in more detail in Gillete *et al.* (*Nat Methods*. 2013 Jan;10(1):28-34), herein incorporated by reference in its entirety for all purposes. Targeted mass spectrometry analysis, including use of K562 single allelic *in vitro* systems, is described in international application publication WO2021092095, herein incorporated by reference for all purposes. Results are shown in **Table 5** demonstrating confirmation of HLA-specific presentation by various KRAS neopeptides.

Table 5 – KRAS Neopeptide Mass Spectrometry

Mutation	HLA	Sequence	Tumors	20x1 K562 Cell Lines	4x4 K562 Cell Lines	8x2 K562 Cell Lines
G12C	A*03:01	VVVGACGVGK	No	Yes	NA	NA
G12C	A*11:01	VVVGACGVGK	No	Yes	Yes	No
G12D	A*03:01	VVGADGVGK	No	Yes	NA	NA
G12D	A*03:01	VVVGADGVGK	No	Yes	NA	NA
G12D	A*11:01	VVGADGVGK	Yes	Yes	Yes	No
G12D	A*11:01	VVVGADGVGK	Yes	Yes	Yes	Yes
G12V	A*02:01	KLVVVGAVGV	No	No	Yes	No
G12V	A*03:01	VVGAVGVGK	No	Yes	NA	NA
G12V	A*03:01	VVVGAVGVGK	Yes	Yes	NA	NA
G12V	A*11:01	VVGAVGVGK	Yes	Yes	Yes	Yes
G12V	A*11:01	VVVGAVGVGK	Yes	Yes	Yes	Yes
G12V	C*01:02	AVGVGKSAL	Yes	NA	NA	NA
Q61H	A*01:01	ILDTAGHEEY	NA	Yes	Yes	Yes

NA = not applicable (either tumor or cell line not available/generated to test)

XIV.F. Iterated KRAS Neopeptide Cassette Evaluation Results

[00451] Vaccine efficacy with various cassettes having repeated (*i.e.*, iterated) neopeptides or a single copy of a given neopeptide were evaluated. Mice engineered to express human HLA-A11:01 for KRAS G12 mutations or human HLA-A01:01 for KRAS Q61H were immunized with 5×10^{10} VP using the ChAdV68 delivery vectors described below and efficacy was assessed by IFN γ ELISpot.

[00452] A series of ChAdV68 delivery vectors were designed to assess efficacy of cassettes featuring either a single copy of KRAS neoepitopes G12C, G12V, G12D, and Q61H (“KRAS 4x1”; cassette = SEQ ID NO: 66), 4 iterations of each of the KRAS neoepitopes G12C, G12V, G12D, and Q61H (“KRAS 4X (4x4)”; cassette = SEQ ID NO: 65), or 4 iterations of only one of the KRAS neoepitopes G12C, G12V, G12D, or Q61H (“KRAS 1x4”).

[00453] Cassettes featuring either a single copy of KRAS neoepitopes (“4x1”) or 4 iterations of each of the KRAS neoepitopes (“4x4”) were compared. As shown in **FIG. 3** and quantified in **Table 6**, vaccination with vectors featuring multiple iterations of the neoepitopes demonstrated increased spot forming colonies (SFC), indicating repetition of epitope-encoding sequences in a cassette led to an increased antigen-specific immune response against KRAS neoepitopes, specifically G12V and G12D neoepitopes, using a ChAdV68 delivery system.

[00454] Cassettes featuring either 4 iterations of each of the KRAS neoepitopes (“4x4”) or 4 iterations of only one of the KRAS (“1x4”) were compared. As shown in **FIG. 4** and quantified in **Table 7**, vaccination with vectors featuring multiple iterations of the neoepitopes in either the 4x4 or 1x4 formats generated immune responses by spot forming colonies (SFC) and were comparable across formats, indicating repetition of epitope-encoding sequences in a cassette led to an increased antigen-specific immune response against KRAS neoepitopes, specifically G12V and G12D neoepitopes, using a ChAdV68 delivery system.

[00455] Immune responses for cassette formats for KRAS Q61H were compared, including cassettes featuring either a single copy of KRAS Q61H with 19 other distinct neoepitopes (“20x1”), 2 iterations of the KRAS Q61H (“8x2”), a single copy of KRAS Q61H together with single copies of KRAS G12 neoepitopes (“4x1”), 4 iterations of the KRAS Q61H together with KRAS G12 neoepitope iterations (“4x4”); or 4 iterations of only KRAS Q61H (“1x4”). As shown in **FIG. 5** and quantified in **Table 8**, while vaccination with a vector a single copy of KRAS Q61H together with single copies of KRAS G12 neoepitopes demonstrated a small improvement over larger cassettes (4x1 vs 20x1 or 8x2), vaccination with vectors featuring multiple iterations of the neoepitopes in either the 4x4 or 1x4 formats generated immune responses by spot forming colonies (SFC) with the 1x4 format demonstrating the most robust response, indicating repetition of epitope-encoding sequences in a cassette as well as shorter formats and/or removal of additional distinct neoepitopes led to an increased antigen-specific immune response against the KRAS neoepitopes Q61H, using a ChAdV68 delivery system.

[00456] Responses following immunization with either a ChAdV68 delivery system or a SAM delivery system we compared. Cassettes featuring either a single copy of KRAS neoepitopes as well as other neoepitopes (“20x1”) or 4 iterations of each of the KRAS neoepitopes (“4x4”) were assessed. As shown in **FIG. 6** and **FIG. 7** and quantified in **Table 9**,

vaccination with vectors featuring multiple iterations of the neoepitopes demonstrated increased spot forming colonies (SFC) for both ChAdV68 and SAM formats, indicating repetition of epitope-encoding sequences in a cassette led to an increased antigen-specific immune response against KRAS neoepitopes, specifically G12V and G12D neoepitopes, using either a ChAdV68 or SAM delivery system suggesting the improved response with cassettes featuring iterations of neoepitopes is agnostic of the vector platform used.

Table 6 – ELISpot data for KRAS Neoepitopes G12V and G12D +/- Repetition

Antigen Cassette	G12V (VVVGAVGGK)		G12D (VVVGADGVGK)	
	Mean	+/- SE	Mean	+/- SE
4x4	5632	237.9	1069	361.4
4x1	1570	833.2	32	20.6

Table 7 – ELISpot data for KRAS Neoepitopes G12V and G12D +/- Repetition

Antigen Cassette	G12V (VVVGAVGGK)		G12D (VVVGADGVGK)*	
	Mean	+/- SE	Mean	+/- SE
4x4	3700	404.9	536	316.1
1x4	3358	310.8	1726	784.7

* Results were not significantly different between formats (p = 0.1896)

Table 8 – ELISpot data for KRAS Neoepitope Q61H +/- Repetition

Antigen Cassette	Q61H (ILDTAGHEEY)	
	Mean	+/- SE
20x1	3	1.3
4x4	554	225
8x2	20	10.8
4x1	262	112.2
1x4	1289	451.7
Naive	1	0.6

Table 9 – ELISpot for KRAS G12V and G12D +/- Repetition in Chad/SAM Formats

	ChAdV68 20x1			SAM 20x1			ChAdV68 4x4			SAM 4x4		
	Mean	SEM	N	Mean	SEM	N	Mean	SEM	N	Mean	SEM	N
G12V pool	13	6	6	33	6	6	4483	767	6	6939	1149	6
G12D pool	5	2	6	6	3	6	297	227	6	560	163	6
G12V (VVGAVGVGK)	4	2	6	12	3	6	3167	846	6	6878	1216	6
G12D (VVVGADGVGK)	4	3	6	8	3	6	356	288	6	564	148	6

Certain Sequences

[00457] Vectors, cassettes, and antibodies referred to herein are described below and referred to by SEQ ID NO.

Tremelimumab VL (SEQ ID NO:16)
Tremelimumab VH (SEQ ID NO:17)
Tremelimumab VH CDR1 (SEQ ID NO:18)
Tremelimumab VH CDR2 (SEQ ID NO:19)
Tremelimumab VH CDR3 (SEQ ID NO:20)
Tremelimumab VL CDR1 (SEQ ID NO:21)
Tremelimumab VL CDR2 (SEQ ID NO:22)
Tremelimumab VL CDR3 (SEQ ID NO:23)
Durvalumab (MEDI4736) VL (SEQ ID NO:24)
MEDI4736 VH (SEQ ID NO:25)
MEDI4736 VH CDR1 (SEQ ID NO:26)
MEDI4736 VH CDR2 (SEQ ID NO:27)
MEDI4736 VH CDR3 (SEQ ID NO:28)
MEDI4736 VL CDR1 (SEQ ID NO:29)
MEDI4736 VL CDR2 (SEQ ID NO:30)
MEDI4736 VL CDR3 (SEQ ID NO:31)
UbA76-25merPDTT 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 [VEE] (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
VEE Delivery Vector (SEQ ID NO:6); VEE genome with nucleotides 7544-11175 deleted [alphavirus structural proteins removed]
TC-83 Delivery Vector(SEQ ID NO:7); TC-83 genome with nucleotides 7544-11175 deleted [alphavirus structural proteins removed]
VEE Production Vector (SEQ ID NO:8); VEE genome with nucleotides 7544-11175 deleted, plus 5' T7-promoter, plus 3' restriction sites
TC-83 Production Vector(SEQ ID NO:9); TC-83 genome with nucleotides 7544-11175 deleted, plus 5' T7-promoter, plus 3' restriction sites
VEE-UbAAY (SEQ ID NO:14); VEE delivery vector with MHC class I mouse tumor epitopes SIINFEKL and AH1-A5 inserted
VEE-Luciferase (SEQ ID NO:15); VEE delivery vector with luciferase gene inserted at 7545
ubiquitin (SEQ ID NO:38)>UbG76 0-228
Ubiquitin A76 (SEQ ID NO:39)>UbA76 0-228
HLA-A2 (MHC class I) signal peptide (SEQ ID NO:40)>MHC SignalPep 0-78
HLA-A2 (MHC class I) Trans Membrane domain (SEQ ID NO:41)>HLA A2 TM Domain 0-201
IgK Leader Seq (SEQ ID NO:42)>IgK Leader Seq 0-60
Human DC-Lamp (SEQ ID NO:43)>HumanDCLAMP 0-3178
Mouse LAMP1 (SEQ ID NO:44)>MouseLamp1 0-1858
Human Lamp1 cDNA (SEQ ID NO:45)>Human Lamp1 0-2339
Tetanus toxoid nucleic acid sequence (SEQ ID NO:46)
Tetanus toxoid amino acid sequence (SEQ ID NO:47)
PADRE nucleotide sequence (SEQ ID NO:48)
PADRE amino acid sequence (SEQ ID NO:49)
WPRES (SEQ ID NO:50)>WPRES 0-593
IRES (SEQ ID NO:51)>eGFP_IRES_SEAP_Insert 1746-2335

ATGAGAATGACGGAATACAAACTTGTGTCGTGGGTGCGGATGGAGTAGGGAAAAGTGCTCTAACA
 ATACAACTCATTGAGATGACAGAGTACAAATTGGTAGTCGTCGGTTCGGTAGGAGTTGGGAAGTCT
 GCACTAACTATTGAGCTCATACAGATGACCGAGTACAAGCTGGTGGTGGTAGGCGCTTTCGGGTGTG
 GGTAAGAGTGCATTAACCATAACAGCTTATACAAGAGACATGTCTGCTAGATATATTAGATACCGCC
 GGGCATGAAGAGTACTCTGCCATGCGAGACCAATACATGCGTATGACAGAGTATAAATTAGTAGTG
 GTTGGGGCGGACGGTGTGGCAAGAGCGCCTTAACCTATACAGTTGATCCAGATGACGGAGTACAAA
 CTGGTTCGTCGTTGGTGCAGTGGGAGTGGGAAAATCTGCGCTGACGATTCAACTAATCCAAGAAACA
 TGTTTACTTGACATCCTCGACACTGCGGGTCACGAGGAGTATTCGGCGATGCGTGATCAATATATGA
 GGATGACTGAGTATAAGTTAGTCGTAGTTGGAGCGGTTCGGTGTGCGGAAAGTCCGCGCTAACCATTC
 AATTGATTCAAAATGACTGAATACAAGCTAGTGGTAGTAGGAGCATGCGGCGTCGGCAAAATCGGCTT
 TAACAATCCAATGATACAGGGACCCGGACCAGGCGCCAAATTTGTTGCTGCTTGGACACTGAAAG
 CTGCTGCTGGGCCCCGACCAGGCCAGTACATCAAGGCCAACTCTAAGTTTATCGGCATCACCGAAT
 TGGGACCTGGACCCGGCTAG

KRAS 4x1 Cassette Nucleotide Sequence (SEQ ID NO: 72)

ATGGCTGGCATGACCGAGTATAAACTAGTAGTTGTGGGAGCGTGTGGTGTAGGCAAGTCGGCACTT
 ACAATTCAGTTGATACAAATGACGGAATATAAGCTCGTAGTAGTCGGAGCAGACGGCGTGGGGAA
 ATCAGCGTTGACTATCCAGTTAATACAGGAACTTGCCTATTAGACATCTTGGATACGGCAGGTCAT
 GAGGAATATTCCGCTATGAGAGATCAGTATATGCGCATGACCGAATACAAGTTGGTTCGTGGTGGG
 GCAGTTGGGGTCGGTAAATCCGCGTTAACGATCCAACCTTATCCAAGGACCCGGACCAGGCGCCAAA
 TTTGTTGCTGCTTGGACACTGAAAGCTGCTGCTGGGCCCCGACCAGGCCAGTACATCAAGGCCAAC
 TCTAAGTTTATCGGCATCACCGAATTGGGACCTGGACCCGGCTAG

KRAS 4x1 + R213L Cassette Nucleotide Sequence (SEQ ID NO: 73)

ATGGCTGGCATGACCGAGTATAAACTAGTAGTTGTGGGAGCGTGTGGTGTAGGCAAGTCGGCACTT
 ACAATTCAGTTGATACAAATGACGGAATATAAGCTCGTAGTAGTCGGAGCAGACGGCGTGGGGAA
 ATCAGCGTTGACTATCCAGTTAATACAGGAACTTGCCTATTAGACATCTTGGATACGGCAGGTCAT
 GAGGAATATTCCGCTATGAGAGATCAGTATATGCGCATGACCGAATACAAGTTGGTTCGTGGTGGG
 GCAGTTGGGGTCGGTAAATCCGCGTTAACGATCCAACCTTATCCAACACTGCGCGTGGAGTATCTGGAT
 GACAGGAACACCTTCCTGCATTCTGTTGTGGTGCCTATGAGCCCCCGAGGTGGGACCCGGACCA
 GCGCCAAATTTGTTGCTGCTTGGACACTGAAAGCTGCTGCTGGGCCCCGACCAGGCCAGTACATC
 AAGGCCAACTCTAAGTTTATCGGCATCACCGAATTGGGACCTGGACCCGGCTAG

KRAS 4x1 + S127Y Cassette Nucleotide Sequence (SEQ ID NO: 74)

ATGGCTGGCATGACCGAGTATAAACTAGTAGTTGTGGGAGCGTGTGGTGTAGGCAAGTCGGCACTT
 ACAATTCAGTTGATACAAATGACGGAATATAAGCTCGTAGTAGTCGGAGCAGACGGCGTGGGGAA
 ATCAGCGTTGACTATCCAGTTAATACAGGAACTTGCCTATTAGACATCTTGGATACGGCAGGTCAT
 GAGGAATATTCCGCTATGAGAGATCAGTATATGCGCATGACCGAATACAAGTTGGTTCGTGGTGGG
 GCAGTTGGGGTCGGTAAATCCGCGTTAACGATCCAACCTTATCCAACACAGCGGCACCGCCAAATCT
 GTCACGTGTACTTACTACCCAGCCTTGAACAAGATGTTCTGTCAATTAGCAAAGGGACCCGGACCA
 GCGCCAAATTTGTTGCTGCTTGGACACTGAAAGCTGCTGCTGGGCCCCGACCAGGCCAGTACATC
 AAGGCCAACTCTAAGTTTATCGGCATCACCGAATTGGGACCTGGACCCGGCTAG

KRAS G12C Neopeptide (SEQ ID NO: 75)

VVVGACGVGK

KRAS G12C Neopeptide (SEQ ID NO: 76)

KLVVVGACGV

KRAS G12D Neopeptide (SEQ ID NO: 77)

VVGADGVGK

KRAS G12D Neopeptide (SEQ ID NO: 78)

VVVGADGVGK

KRAS G12V Neopeptide (SEQ ID NO: 79)

VVGAVGVGK

KRAS G12V Neopeptide (SEQ ID NO: 80)

AVGVGKSAL

KRAS G12V Neopeptide (SEQ ID NO: 81)

VVVGAVGVGK

KRAS Q61H Neopeptide (SEQ ID NO: 82)

ILDTAGHEEY

References

1. Desrichard, A., Snyder, A. & Chan, T. A. Cancer Neoantigens and Applications for Immunotherapy. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* (2015). doi:10.1158/1078-0432.CCR-14-3175
2. Schumacher, T. N. & Schreiber, R. D. Neoantigens in cancer immunotherapy. *Science* **348**, 69–74 (2015).
3. Gubin, M. M., Artyomov, M. N., Mardis, E. R. & Schreiber, R. D. Tumor neoantigens: building a framework for personalized cancer immunotherapy. *J. Clin. Invest.* **125**, 3413–3421 (2015).
4. Rizvi, N. A. *et al.* Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science* **348**, 124–128 (2015).
5. Snyder, A. *et al.* Genetic basis for clinical response to CTLA-4 blockade in melanoma. *N. Engl. J. Med.* **371**, 2189–2199 (2014).
6. Carreno, B. M. *et al.* Cancer immunotherapy. A dendritic cell vaccine increases the breadth and diversity of melanoma neoantigen-specific T cells. *Science* **348**, 803–808 (2015).
7. Tran, E. *et al.* Cancer immunotherapy based on mutation-specific CD4⁺ T cells in a patient with epithelial cancer. *Science* **344**, 641–645 (2014).
8. Hacothen, N. & Wu, C. J.-Y. United States Patent Application: 20110293637 - COMPOSITIONS AND METHODS OF IDENTIFYING TUMOR SPECIFIC NEOANTIGENS. (A1). at <<http://appft1.uspto.gov/netacgi/nph-Parser?Sect1=PTO1&Sect2=HITOFF&d=PG01&p=1&u=/netahtml/PTO/srchnum.html&r=1&f=G&l=50&s1=20110293637.PG NR.>>>
9. Lundegaard, C., Hoof, I., Lund, O. & Nielsen, M. State of the art and challenges in sequence based T-cell epitope prediction. *Immunome Res.* **6 Suppl 2**, S3 (2010).
10. Yadav, M. *et al.* Predicting immunogenic tumour mutations by combining mass spectrometry and exome sequencing. *Nature* **515**, 572–576 (2014).
11. Bassani-Sternberg, M., Pletscher-Frankild, S., Jensen, L. J. & Mann, M. Mass spectrometry of human leukocyte antigen class I peptidomes reveals strong effects of protein abundance and turnover on antigen presentation. *Mol. Cell. Proteomics MCP* **14**, 658–673 (2015).
12. Van Allen, E. M. *et al.* Genomic correlates of response to CTLA-4 blockade in metastatic melanoma. *Science* **350**, 207–211 (2015).
13. Yoshida, K. & Ogawa, S. Splicing factor mutations and cancer. *Wiley Interdiscip. Rev. RNA* **5**, 445–459 (2014).
14. Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma. *Nature* **511**, 543–550 (2014).
15. Rajasagi, M. *et al.* Systematic identification of personal tumor-specific neoantigens in chronic lymphocytic leukemia. *Blood* **124**, 453–462 (2014).
16. Downing, S. R. *et al.* United States Patent Application: 0120208706 - OPTIMIZATION OF MULTIGENE ANALYSIS OF TUMOR SAMPLES. (A1). at <<http://appft1.uspto.gov/netacgi/nph-Parser?Sect1=PTO1&Sect2=HITOFF&d=PG01&p=1&u=/netahtml/PTO/srchnum.html&r=1&f=G&l=50&s1=20120208706.PG NR.>>>
17. Target Capture for NextGen Sequencing - IDT. at <<http://www.idtdna.com/pages/products/nextgen/target-capture>>
18. Shukla, S. A. *et al.* Comprehensive analysis of cancer-associated somatic mutations in class I HLA genes. *Nat. Biotechnol.* **33**, 1152–1158 (2015).
19. Cieslik, M. *et al.* The use of exome capture RNA-seq for highly degraded RNA with application to clinical cancer sequencing. *Genome Res.* **25**, 1372–1381 (2015).
20. Bodini, M. *et al.* The hidden genomic landscape of acute myeloid leukemia: subclonal structure revealed by undetected mutations. *Blood* **125**, 600–605 (2015).

21. Saunders, C. T. *et al.* Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. *Bioinforma. Oxf. Engl.* **28**, 1811–1817 (2012).
22. Cibulskis, K. *et al.* Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat. Biotechnol.* **31**, 213–219 (2013).
23. Wilkerson, M. D. *et al.* Integrated RNA and DNA sequencing improves mutation detection in low purity tumors. *Nucleic Acids Res.* **42**, e107 (2014).
24. Mose, L. E., Wilkerson, M. D., Hayes, D. N., Perou, C. M. & Parker, J. S. ABRA: improved coding indel detection via assembly-based realignment. *Bioinforma. Oxf. Engl.* **30**, 2813–2815 (2014).
25. Ye, K., Schulz, M. H., Long, Q., Apweiler, R. & Ning, Z. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinforma. Oxf. Engl.* **25**, 2865–2871 (2009).
26. Lam, H. Y. K. *et al.* Nucleotide-resolution analysis of structural variants using BreakSeq and a breakpoint library. *Nat. Biotechnol.* **28**, 47–55 (2010).
27. Frampton, G. M. *et al.* Development and validation of a clinical cancer genomic profiling test based on massively parallel DNA sequencing. *Nat. Biotechnol.* **31**, 1023–1031 (2013).
28. Boegel, S. *et al.* HLA typing from RNA-Seq sequence reads. *Genome Med.* **4**, 102 (2012).
29. Liu, C. *et al.* ATHLATES: accurate typing of human leukocyte antigen through exome sequencing. *Nucleic Acids Res.* **41**, e142 (2013).
30. Mayor, N. P. *et al.* HLA Typing for the Next Generation. *PloS One* **10**, e0127153 (2015).
31. Roy, C. K., Olson, S., Graveley, B. R., Zamore, P. D. & Moore, M. J. Assessing long-distance RNA sequence connectivity via RNA-templated DNA-DNA ligation. *eLife* **4**, (2015).
32. Song, L. & Florea, L. CLASS: constrained transcript assembly of RNA-seq reads. *BMC Bioinformatics* **14 Suppl 5**, S14 (2013).
33. Maretty, L., Sibbesen, J. A. & Krogh, A. Bayesian transcriptome assembly. *Genome Biol.* **15**, 501 (2014).
34. Pertea, M. *et al.* StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **33**, 290–295 (2015).
35. Roberts, A., Pimentel, H., Trapnell, C. & Pachter, L. Identification of novel transcripts in annotated genomes using RNA-Seq. *Bioinforma. Oxf. Engl.* (2011). doi:10.1093/bioinformatics/btr355
36. Vitting-Seerup, K., Porse, B. T., Sandelin, A. & Waage, J. spliceR: an R package for classification of alternative splicing and prediction of coding potential from RNA-seq data. *BMC Bioinformatics* **15**, 81 (2014).
37. Rivas, M. A. *et al.* Human genomics. Effect of predicted protein-truncating genetic variants on the human transcriptome. *Science* **348**, 666–669 (2015).
38. Skelly, D. A., Johansson, M., Madeoy, J., Wakefield, J. & Akey, J. M. A powerful and flexible statistical framework for testing hypotheses of allele-specific gene expression from RNA-seq data. *Genome Res.* **21**, 1728–1737 (2011).
39. Anders, S., Pyl, P. T. & Huber, W. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinforma. Oxf. Engl.* **31**, 166–169 (2015).
40. Furney, S. J. *et al.* SF3B1 mutations are associated with alternative splicing in uveal melanoma. *Cancer Discov.* (2013). doi:10.1158/2159-8290.CD-13-0330
41. Zhou, Q. *et al.* A chemical genetics approach for the functional assessment of novel cancer genes. *Cancer Res.* (2015). doi:10.1158/0008-5472.CAN-14-2930
42. Maguire, S. L. *et al.* SF3B1 mutations constitute a novel therapeutic target in breast cancer. *J. Pathol.* **235**, 571–580 (2015).
43. Carithers, L. J. *et al.* A Novel Approach to High-Quality Postmortem Tissue Procurement: The GTEx Project. *Biopreservation Biobanking* **13**, 311–319 (2015).
44. Xu, G. *et al.* RNA CoMPASS: a dual approach for pathogen and host transcriptome analysis of RNA-seq datasets. *PloS One* **9**, e89445 (2014).

45. Andreatta, M. & Nielsen, M. Gapped sequence alignment using artificial neural networks: application to the MHC class I system. *Bioinforma. Oxf. Engl.* (2015). doi:10.1093/bioinformatics/btv639
46. Jørgensen, K. W., Rasmussen, M., Buus, S. & Nielsen, M. NetMHCstab - predicting stability of peptide-MHC-I complexes; impacts for cytotoxic T lymphocyte epitope discovery. *Immunology* **141**, 18–26 (2014).
47. Larsen, M. V. *et al.* An integrative approach to CTL epitope prediction: a combined algorithm integrating MHC class I binding, TAP transport efficiency, and proteasomal cleavage predictions. *Eur. J. Immunol.* **35**, 2295–2303 (2005).
48. Nielsen, M., Lundegaard, C., Lund, O. & Keşmir, C. The role of the proteasome in generating cytotoxic T-cell epitopes: insights obtained from improved predictions of proteasomal cleavage. *Immunogenetics* **57**, 33–41 (2005).
49. Boisvert, F.-M. *et al.* A Quantitative Spatial Proteomics Analysis of Proteome Turnover in Human Cells. *Mol. Cell. Proteomics* **11**, M111.011429–M111.011429 (2012).
50. Duan, F. *et al.* Genomic and bioinformatic profiling of mutational neoepitopes reveals new rules to predict anticancer immunogenicity. *J. Exp. Med.* **211**, 2231–2248 (2014).
51. Janeway's Immunobiology: 9780815345312: Medicine & Health Science Books @ Amazon.com. at <<http://www.amazon.com/Janeways-Immunobiology-Kenneth-Murphy/dp/0815345313>>
52. Calis, J. J. A. *et al.* Properties of MHC Class I Presented Peptides That Enhance Immunogenicity. *PLoS Comput. Biol.* **9**, e1003266 (2013).
53. Zhang, J. *et al.* Intratumor heterogeneity in localized lung adenocarcinomas delineated by multiregion sequencing. *Science* **346**, 256–259 (2014)
54. Walter, M. J. *et al.* Clonal architecture of secondary acute myeloid leukemia. *N. Engl. J. Med.* 366, 1090–1098 (2012).
55. Hunt DF, Henderson RA, Shabanowitz J, Sakaguchi K, Michel H, Sevilir N, Cox AL, Appella E, Engelhard VH. Characterization of peptides bound to the class I MHC molecule HLA-A2.1 by mass spectrometry. *Science* 1992. 255: 1261-1263.
56. Zarling AL, Polefrone JM, Evans AM, Mikesh LM, Shabanowitz J, Lewis ST, Engelhard VH, Hunt DF. Identification of class I MHC-associated phosphopeptides as targets for cancer immunotherapy. *Proc Natl Acad Sci U S A.* 2006 Oct 3;103(40):14889-94.
57. Bassani-Sternberg M, Pletscher-Frankild S, Jensen LJ, Mann M. Mass spectrometry of human leukocyte antigen class I peptidomes reveals strong effects of protein abundance and turnover on antigen presentation. *Mol Cell Proteomics.* 2015 Mar;14(3):658-73. doi: 10.1074/mcp.M114.042812.
58. Abelin JG, Trantham PD, Penny SA, Patterson AM, Ward ST, Hildebrand WH, Cobbold M, Bai DL, Shabanowitz J, Hunt DF. Complementary IMAC enrichment methods for HLA-associated phosphopeptide identification by mass spectrometry. *Nat Protoc.* 2015 Sep;10(9):1308-18. doi: 10.1038/nprot.2015.086. Epub 2015 Aug 6
59. Barnstable CJ, Bodmer WF, Brown G, Galfre G, Milstein C, Williams AF, Ziegler A. Production of monoclonal antibodies to group A erythrocytes, HLA and other human cell surface antigens-new tools for genetic analysis. *Cell.* 1978 May;14(1):9-20.
60. Goldman JM, Hibbin J, Kearney L, Orchard K, Th'ng KH. HLA-DR monoclonal antibodies inhibit the proliferation of normal and chronic granulocytic leukaemia myeloid progenitor cells. *Br J Haematol.* 1982 Nov;52(3):411-20.
61. Eng JK, Jahan TA, Hoopmann MR. Comet: an open-source MS/MS sequence database search tool. *Proteomics.* 2013 Jan;13(1):22-4. doi: 10.1002/pmic.201200439. Epub 2012 Dec 4.
62. Eng JK, Hoopmann MR, Jahan TA, Egertson JD, Noble WS, MacCoss MJ. A deeper look into Comet--implementation and features. *J Am Soc Mass Spectrom.* 2015 Nov;26(11):1865-74. doi: 10.1007/s13361-015-1179-x. Epub 2015 Jun 27.

63. Lukas Käll, Jesse Canterbury, Jason Weston, William Stafford Noble and Michael J. MacCoss. Semi-supervised learning for peptide identification from shotgun proteomics datasets. *Nature Methods* 4:923 – 925, November 2007
64. Lukas Käll, John D. Storey, Michael J. MacCoss and William Stafford Noble. Assigning confidence measures to peptides identified by tandem mass spectrometry. *Journal of Proteome Research*, 7(1):29-34, January 2008
65. Lukas Käll, John D. Storey and William Stafford Noble. Nonparametric estimation of posterior error probabilities associated with peptides identified by tandem mass spectrometry. *Bioinformatics*, 24(16):i42-i48, August 2008
66. Kinney RM, BJ Johnson , VL Brown , DW Trent. Nucleotide Sequence of the 26 S mRNA of the Virulent Trinidad Donkey Strain of Venezuelan Equine Encephalitis Virus and Deduced Sequence of the Encoded Structural Proteins. *Virology* 152 (2), 400-413. 1986 Jul 30.
67. Jill E Slansky, Frédérique M Rattis, Lisa F Boyd, Tarek Fahmy, Elizabeth M Jaffee, Jonathan P Schneck, David H Margulies, Drew M Pardoll. Enhanced Antigen-Specific Antitumor Immunity with Altered Peptide Ligands that Stabilize the MHC-Peptide-TCR Complex. *Immunity*, Volume 13, Issue 4, 1 October 2000, Pages 529-538.
68. A Y Huang, P H Gulden, A S Woods, M C Thomas, C D Tong, W Wang, V H Engelhard, G Pasternack, R Cotter, D Hunt, D M Pardoll, and E M Jaffee. The immunodominant major histocompatibility complex class I-restricted antigen of a murine colon tumor derives from an endogenous retroviral gene product. *Proc Natl Acad Sci U S A.*; 93(18): 9730–9735, 1996 Sep 3.
69. JOHNSON, BARBARA J. B., RICHARD M. KINNEY, CRYSTLE L. KOST AND DENNIS W. TRENT. Molecular Determinants of Alphavirus Neurovirulence: Nucleotide and Deduced Protein Sequence Changes during Attenuation of Venezuelan Equine Encephalitis Virus. *J Gen Virol* 67:1951-1960, 1986.
70. Aarnoudse, C.A., Krüse, M., Konopitzky, R., Brouwenstijn, N., and Schrier, P.I. (2002). TCR reconstitution in Jurkat reporter cells facilitates the identification of novel tumor antigens by cDNA expression cloning. *Int J Cancer* 99, 7–13.
71. Alexander, J., Sidney, J., Southwood, S., Ruppert, J., Oseroff, C., Maewal, A., Snoke, K., Serra, H.M., Kubo, R.T., and Sette, A. (1994). Development of high potency universal DR-restricted helper epitopes by modification of high affinity DR-blocking peptides. *Immunity* 1, 751–761.
72. Banu, N., Chia, A., Ho, Z.Z., Garcia, A.T., Paravasivam, K., Grotenbreg, G.M., Bertoletti, A., and Gehring, A.J. (2014). Building and optimizing a virus-specific T cell receptor library for targeted immunotherapy in viral infections. *Scientific Reports* 4, 4166.
73. Cornet, S., Miconnet, I., Menez, J., Lemonnier, F., and Kosmatopoulos, K. (2006). Optimal organization of a polypeptide-based candidate cancer vaccine composed of cryptic tumor peptides with enhanced immunogenicity. *Vaccine* 24, 2102–2109.
74. Depla, E., van der Aa, A., Livingston, B.D., Crimi, C., Allosery, K., de Brabandere, V., Krakover, J., Murthy, S., Huang, M., Power, S., et al. (2008). Rational design of a multiepitope vaccine encoding T-lymphocyte epitopes for treatment of chronic hepatitis B virus infections. *Journal of Virology* 82, 435–450.
75. Ishioka, G.Y., Fikes, J., Hermanson, G., Livingston, B., Crimi, C., Qin, M., del Guercio, M.F., Oseroff, C., Dahlberg, C., Alexander, J., et al. (1999). Utilization of MHC class I transgenic mice for development of minigene DNA vaccines encoding multiple HLA-restricted CTL epitopes. *J Immunol* 162, 3915–3925.
76. Janetzki, S., Price, L., Schroeder, H., Britten, C.M., Welters, M.J.P., and Hoos, A. (2015). Guidelines for the automated evaluation of Elispot assays. *Nat Protoc* 10, 1098–1115.
77. Lyons, G.E., Moore, T., Brasic, N., Li, M., Roszkowski, J.J., and Nishimura, M.I. (2006). Influence of human CD8 on antigen recognition by T-cell receptor-transduced cells. *Cancer Res* 66, 11455–11461.
78. Nagai, K., Ochi, T., Fujiwara, H., An, J., Shirakata, T., Mineno, J., Kuzushima, K., Shiku, H., Melenhorst, J.J., Gostick, E., et al. (2012). Aurora kinase A-specific T-cell receptor gene

- transfer redirects T lymphocytes to display effective antileukemia reactivity. *Blood* 119, 368–376.
79. Panina-Bordignon, P., Tan, A., Termijtellen, A., Demotz, S., Corradin, G., and Lanzavecchia, A. (1989). Universally immunogenic T cell epitopes: promiscuous binding to human MHC class II and promiscuous recognition by T cells. *Eur J Immunol* 19, 2237–2242.
80. Vitiello, A., Marchesini, D., Furze, J., Sherman, L.A., and Chesnut, R.W. (1991). Analysis of the HLA-restricted influenza-specific cytotoxic T lymphocyte response in transgenic mice carrying a chimeric human-mouse class I major histocompatibility complex. *J Exp Med* 173, 1007–1015.
81. Yachi, P.P., Ampudia, J., Zal, T., and Gascoigne, N.R.J. (2006). Altered peptide ligands induce delayed CD8-T cell receptor interaction--a role for CD8 in distinguishing antigen quality. *Immunity* 25, 203–211.
82. Pushko P, Parker M, Ludwig GV, Davis NL, Johnston RE, Smith JF. Replicon-helper systems from attenuated Venezuelan equine encephalitis virus: expression of heterologous genes in vitro and immunization against heterologous pathogens in vivo. *Virology*. 1997 Dec 22;239(2):389-401.
83. Strauss, JH and E G Strauss. The alphaviruses: gene expression, replication, and evolution. *Microbiol Rev*. 1994 Sep; 58(3): 491–562.
84. Rhône C, Ehrenguber MU, Grandgirard D. Alphaviral cytotoxicity and its implication in vector development. *Exp Physiol*. 2005 Jan;90(1):45-52. Epub 2004 Nov 12.
85. Riley, Michael K. II, and Wilfred Vermerris. Recent Advances in Nanomaterials for Gene Delivery—A Review. *Nanomaterials* 2017, 7(5), 94.
86. Frolov I, Hardy R, Rice CM. Cis-acting RNA elements at the 5' end of Sindbis virus genome RNA regulate minus- and plus-strand RNA synthesis. *RNA*. 2001 Nov;7(11):1638-51.
87. Jose J, Snyder JE, Kuhn RJ. A structural and functional perspective of alphavirus replication and assembly. *Future Microbiol*. 2009 Sep;4(7):837-56.
88. Bo Li and C. olin N. Dewey. RSEM: accurate transcript quantification from RNA-Seq data with or without a referenfe genome. *BMC Bioinformatics*, 12:323, August 2011
89. Hillary Pearson, Tariq Daouda, Diana Paola Granados, Chantal Durette, Eric Bonneil, Mathieu Courcelles, Anja Rodenbrock, Jean-Philippe Laverdure, Caroline Côté, Sylvie Mader, Sébastien Lemieux, Pierre Thibault, and Claude Perreault. MHC class I-associated peptides derive from selective regions of the human genome. *The Journal of Clinical Investigation*, 2016, 90.
90. Juliane Liepe, Fabio Marino, John Sidney, Anita Jeko, Daniel E. Bunting, Alessandro Sette, Peter M. Klotzel, Michael P. H. Stumpf, Albert J. R. Heck, Michele Mishto. A large fraction of HLA class I ligands are proteasome-generated spliced peptides. *Science*, 21, October 2016.
91. Mommen GP., Marino, F., Meiring HD., Poelen, MC., van Gaans-van den Brink, JA., Mohammed S., Heck AJ., and van Els CA. Sampling From the Proteome to the Human Leukocyte Antigen-DR (HLA-DR) Ligandome Proceeds Via High Specificity. *Mol Cell Proteomics* 15(4): 1412-1423, April 2016.
92. Sebastian Kreiter, Mathias Vormehr, Niels van de Roemer, Mustafa Diken, Martin Löwer, Jan Diekmann, Sebastian Boegel, Barbara Schrörs, Fulvia Vascotto, John C. Castle, Arbel D. Tadmor, Stephen P. Schoenberger, Christoph Huber, Özlem Türeci, and Ugur Sahin. Mutant MHC class II epitopes drive therapeutic immune responses to cancer. *Nature* 520, 692-696, April 2015.
93. Tran E., Turcotte S., Gros A., Robbins P.F., Lu Y.C., Dudley M.E., Wunderlich J.R., Somerville R.P., Hogan K., Hinrichs C.S., Parkhurst M.R., Yang J.C., Rosenberg S.A. Cancer immunotherapy based on mutation-specific CD4+ T cells in a patient with epithelial cancer. *Science* 344(6184) 641-645, May 2014.
94. Andreatta M., Karosiene E., Rasmussen M., Stryhn A., Buus S., Nielsen M. Accurate pan-specific prediction of peptide-MHC class II binding affinity with improved binding core identification. *Immunogenetics* 67(11-12) 641-650, November 2015.

95. Nielsen, M., Lund, O. NN-align. An artificial neural network-based alignment algorithm for MHC class II peptide binding prediction. *BMC Bioinformatics* 10:296, September 2009.
96. Nielsen, M., Lundegaard, C., Lund, O. Prediction of MHC class II binding affinity using SMM-align, a novel stabilization matrix alignment method. *BMC Bioinformatics* 8:238, July 2007.
97. Zhang, J., et al. PEAKS DB: de novo sequencing assisted database search for sensitive and accurate peptide identification. *Molecular & Cellular Proteomics*. 11(4):1-8. 1/2/2012.
98. Jensen, Kamilla Kjaergaard, et al. "Improved Methods for Predicting Peptide Binding Affinity to MHC Class II Molecules." *Immunology*, 2018, doi:10.1111/imm.12889.
99. Carter, S.L., Cibulskis, K., Helman, E., McKenna, A., Shen, H., Zack, T., Laird, P.W., Onofrio, R.C., Winckler, W., Weir, B.A., et al. (2012). Absolute quantification of somatic DNA alterations in human cancer. *Nat. Biotechnol.* 30, 413–421
100. McGranahan, N., Rosenthal, R., Hiley, C.T., Rowan, A.J., Watkins, T.B.K., Wilson, G.A., Birkbak, N.J., Veeriah, S., Van Loo, P., Herrero, J., et al. (2017). Allele-Specific HLA Loss and Immune Escape in Lung Cancer Evolution. *Cell* 171, 1259–1271.e11.
101. Shukla, S.A., Rooney, M.S., Rajasagi, M., Tiao, G., Dixon, P.M., Lawrence, M.S., Stevens, J., Lane, W.J., Dellagatta, J.L., Steelman, S., et al. (2015). Comprehensive analysis of cancer-associated somatic mutations in class I HLA genes. *Nat. Biotechnol.* 33, 1152–1158.
102. Van Loo, P., Nordgard, S.H., Lingjærde, O.C., Russnes, H.G., Rye, I.H., Sun, W., Weigman, V.J., Marynen, P., Zetterberg, A., Naume, B., et al. (2010). Allele-specific copy number analysis of tumors. *Proc. Natl. Acad. Sci. U. S. A.* 107, 16910–16915.
103. Van Loo, P., Nordgard, S.H., Lingjærde, O.C., Russnes, H.G., Rye, I.H., Sun, W., Weigman, V.J., Marynen, P., Zetterberg, A., Naume, B., et al. (2010). Allele-specific copy number analysis of tumors. *Proc. Natl. Acad. Sci. U. S. A.* 107, 16910–16915.
104. MacLean, B., Tomazela, D.M., Shulman, N., Chambers, M., Finney, G.L., Frewen, B., Kern, R., Tabb, D.L., Liebler, D.C., MacCoss, M.J., (2010). Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* 26(7),966-8.

CLAIMS

What is claimed is:

1. An antigen-encoding cassette, or a polypeptide sequence encoded by the cassette, wherein the antigen-encoding cassette comprises at least one antigen-encoding nucleic acid sequence described, from 5' to 3', by the formula:

$$(E_x-(E^N)_y)_z$$
 wherein E represents a nucleotide sequence comprising a distinct epitope-encoding nucleic acid sequence,
 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, and
 at least one of the distinct epitope-encoding nucleic acid sequences comprising the at least two iterations encodes a KRAS-associated MHC class I neoepitope.
2. The composition of claim 1, wherein the antigen-encoding cassette encodes at least 4 iterations of each of the amino acid sequences VVGACGVGK (SEQ ID NO: 75), VVGADGVGK (SEQ ID NO: 78), VVGAVGVGK (SEQ ID NO: 79), and ILDTAGHEEY (SEQ ID NO: 82).
3. The composition of claim 1, wherein at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, or at least 8 of the distinct epitope-encoding nucleic acid sequences comprising the at least two iterations encode distinct KRAS-associated MHC class I neoepitopes.
4. The composition of claim 1, wherein each of the distinct epitope-encoding nucleic acid sequences comprising the at least two iterations encode distinct KRAS-associated MHC class I neoepitopes.
5. The composition of any one of claims 1-4, wherein one or more of the epitope-encoding nucleic acid sequences encoding the KRAS-associated MHC class I neoepitopes comprises at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, or at least 8 iterations.

6. The composition of any one of claims 1-4, wherein each of the nucleic acid sequences encoding the KRAS-associated MHC class I neoepitopes comprises at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, or at least 8 iterations.
7. The composition of any one of claims 1-4, wherein one or more of the nucleic acid sequences encoding the distinct KRAS-associated MHC class I neoepitopes comprises at least 4 iterations.
8. The composition of any one of claims 1-4, wherein each of the nucleic acid sequences encoding the distinct KRAS-associated MHC class I neoepitopes comprises at least 4 iterations.
9. The composition of any one of claims 1-8, wherein one or more of the distinct KRAS-associated MHC class I neoepitopes independently comprises a KRAS G12C mutation, a KRAS G12V mutation, a KRAS G12D mutation, or a KRAS Q61H mutation.
10. The composition of any one of claims 1-9, wherein each E or E^N independently comprises a nucleotide sequence described, from 5' to 3', by the formula (L5_b-N_c-L3_d),
wherein 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.
11. The composition of claim 10, wherein
each N encodes an 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 epitope, and wherein the 5' linker sequence encodes a peptide that is at least 2 amino acids in length, and
L3 is a native 3' linker sequence that encodes a native C-terminal amino acid sequence of the epitope, and wherein the 3' linker sequence encodes a peptide that is at least 2 amino acids in length.
12. The composition of any one of claims 1-11, wherein each E and E^N encodes an epitope at least 7 amino acids in length.
13. The composition of any one of claims 1-11, wherein each E and E^N encodes an epitope 7-15 amino acids in length.
14. The composition of any one of claims 1-13, wherein each E and E^N is a nucleotide sequence at least 21 nucleotides in length.
15. The composition of any one of claims 1-13, wherein each E and E^N is a nucleotide sequence 75 nucleotides in length.
16. A composition for delivery of an antigen expression system, comprising:

the antigen expression system, wherein the antigen expression system comprises one or more vectors,

the one or more vectors comprising:

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

(i) at least one promoter nucleotide sequence, and

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

(b) a cassette, wherein the cassette comprises:

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

(I) an epitope-encoding nucleic acid sequence encoding a KRAS-associated MHC class I neoepitope, and

wherein each of the epitope-encoding nucleic acid sequences comprises;

(A) optionally, a 5' linker sequence, and

(B) optionally, a 3' linker sequence;

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

(iii) optionally, at least one MHC class II epitope-encoding nucleic acid sequence;

(iv) optionally, at least one nucleic acid sequence encoding a GPGPG amino acid linker sequence (SEQ ID NO:56); and

(v) optionally, at least one second poly(A) sequence, wherein the second poly(A) sequence is a native poly(A) sequence or an exogenous poly(A) sequence to the vector backbone,

wherein if the second promoter nucleotide sequence is absent, the antigen-encoding nucleic acid sequence is operably linked to the at least one promoter nucleotide sequence, and

wherein the at least one antigen-encoding nucleic acid sequence comprises at least two iterations of the epitope-encoding nucleic acid sequence encoding the KRAS-associated MHC class I neoepitope.

17. A composition for delivery of an antigen expression system, comprising:
the antigen expression system, wherein the antigen expression system comprises one or more vectors,

the one or more vectors comprising:

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

- (i) at least one promoter nucleotide sequence, and
- (ii) at least one polyadenylation (poly(A)) sequence; and

(b) a cassette, wherein the cassette comprises:

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

- (I) at least 2, 3, 4, 5, 6, 7, 8, 9, or 10 distinct epitope-encoding nucleic acid sequences linearly linked to each other, wherein at least one of the distinct epitope-encoding nucleic acid sequences encodes a KRAS-associated MHC class I neoepitope, and

wherein each of the epitope-encoding nucleic acid sequences comprises;

(A) optionally, a 5' linker sequence, and

(B) optionally, a 3' linker sequence;

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

(iii) optionally, at least one MHC class II epitope-encoding nucleic acid sequence;

(iv) optionally, at least one nucleic acid sequence encoding a GPGPG amino acid linker sequence (SEQ ID NO:56); and

(v) optionally, at least one second poly(A) sequence, wherein the second poly(A) sequence is a native poly(A) sequence or an exogenous poly(A) sequence to the vector backbone,

wherein if the second promoter nucleotide sequence is absent, the antigen-encoding nucleic acid sequence is operably linked to the at least one promoter nucleotide sequence, and

wherein the at least one antigen-encoding nucleic acid sequence comprises at least two repeats of at least one of the distinct epitope-encoding nucleic acid sequences encoding the KRAS-associated MHC class I neoepitope.

18. The composition of claim 17, wherein the at least one antigen-encoding nucleic acid sequence comprises at least two distinct epitope-encoding nucleic acid sequences each encoding a distinct KRAS-associated MHC class I neoepitope.

19. A composition for delivery of an antigen expression system, comprising:
the antigen expression system,
wherein the antigen expression system comprises one or more vectors,
the one or more vectors comprising:
- (a) a vector backbone, wherein the vector backbone comprises a chimpanzee adenovirus vector, optionally wherein the chimpanzee adenovirus vector is a ChAdV68 vector, or an alphavirus vector, optionally wherein the alphavirus vector is a Venezuelan equine encephalitis virus vector; and
 - (b) a cassette, optionally wherein the cassette is integrated between a native promoter nucleotide sequence native to the vector backbone and a poly(A) sequence, optionally wherein the poly(A) sequence is native to the vector backbone, wherein the cassette comprises:
 - (i) at least one antigen-encoding nucleic acid sequence, comprising:
 - (I) an epitope-encoding nucleic acid sequence encoding a KRAS-associated MHC class I neoepitope, optionally comprising at least two distinct epitope-encoding nucleic acid sequences linearly linked to each other, each epitope-encoding nucleic acid sequence optionally comprising:
 - (A) a MHC class I epitope encoding nucleic acid sequence, wherein the MHC class I epitope encoding nucleic acid sequence encodes a MHC class I epitope 7-15 amino acids in length,
 - (B) a 5' linker sequence, wherein the 5' linker sequence encodes a native N-terminal amino acid sequence of the MHC class I epitope, and wherein the 5' linker sequence encodes a peptide that is at least 2 amino acids in length,
 - (C) a 3' linker sequence, wherein the 3' linker sequence encodes a native C-terminal acid sequence of the MHC class I epitope, and wherein the 3' linker sequence encodes a peptide that is at least 2 amino acids in length, and

wherein the cassette is operably linked to the native promoter nucleotide sequence, wherein each of the epitope-encoding nucleic acid sequences encodes a polypeptide that is between 13 and 25 amino acids in length, and wherein each 3' end of each epitope-encoding nucleic acid sequence is linked to the 5' end of the following epitope-encoding nucleic acid sequence with the exception of the final epitope-encoding nucleic acid sequence in the cassette; and

(ii) at least two MHC class II epitope-encoding nucleic acid sequences comprising:

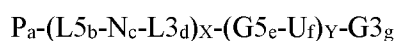
- (I) a PADRE MHC class II sequence (SEQ ID NO:48),
- (II) a Tetanus toxoid MHC class II sequence (SEQ ID NO:46),
- (III) a first nucleic acid sequence encoding a GP GPG amino acid linker sequence linking the PADRE MHC class II sequence and the Tetanus toxoid MHC class II sequence,
- (IV) a second nucleic acid sequence encoding a GP GPG amino acid linker sequence linking the 5' end of the at least two MHC class II epitope-encoding nucleic acid sequences to the epitope-encoding nucleic acid sequences,
- (V) optionally, a third nucleic acid sequence encoding a GP GPG amino acid linker sequence at the 3' end of the at least two MHC class II epitope-encoding nucleic acid sequences;

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

wherein if the second promoter nucleotide sequence is absent, the antigen-encoding nucleic acid sequence is operably linked to the native promoter nucleotide sequence, and

wherein the at least one antigen-encoding nucleic acid sequence comprises at least two iterations of the epitope-encoding nucleic acid sequence encoding the KRAS-associated MHC class I neoepitope.

20. The composition of any one of claims 16-18, wherein an ordered sequence of each element of the cassette 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 distinct epitope-encoding nucleic acid sequences, 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 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.

21. The composition of claim 20, wherein for each X the corresponding N_c is a distinct epitope-encoding nucleic acid sequence, except for the N_c corresponding to the at least two repeats of the distinct epitope-encoding nucleic acid sequence.
22. The composition of claim 20 or 21, wherein for each Y the corresponding U_f is a distinct MHC class II epitope-encoding nucleic acid sequence.
23. The composition of any one of claims 20-22, wherein

$$a = 0, b = 1, d = 1, e = 1, g = 1, h = 1, X = 16, Y = 2,$$

the at least one promoter nucleotide sequence is a single native promoter nucleotide sequence native to the vector backbone,

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

each N encodes an 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 epitope, and wherein the 5' linker sequence encodes a peptide that is at least 2 amino acids in length,

L3 is a native 3' linker sequence that encodes a native C-terminal amino acid sequence of the epitope, and wherein the 3' linker sequence encodes a peptide that is at least 2 amino acids in length,

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

the vector backbone comprises a chimpanzee adenovirus vector, optionally wherein the chimpanzee adenovirus vector is a ChAdV68 vector, or an alphavirus vector, optionally wherein the alphavirus vector is a Venezuelan equine encephalitis virus vector, optionally wherein the native promoter nucleotide sequence is a subgenomic promoter when the vector backbone comprises an alphavirus vector, and each of the MHC class II epitope-encoding nucleic acid sequences encodes a polypeptide that is between 13 and 25 amino acids in length.

24. The composition of any one of claims 16-23, wherein the at least two iterations is at least 3, at least 4, at least 5, at least 6, at least 7, or at least 8 iterations.
25. The composition of any one of claims 16-23, wherein the at least two iterations is at least 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, or at least 20 iterations .
26. The composition of any one of claims 16-23, wherein the at least two iterations is between 2-3, between 2-4, between 2-5, between 2-6, between 2-7 repeats, or between 2-8 iterations.
27. The composition of any one of claims 16-23, wherein the at least two iterations is 7 iterations or less, 6 iterations or less, 5 iterations or less, 4 iterations or less, or 3 iterations or less.
28. The composition of any one of claims 16-27, wherein the at least one antigen-encoding nucleic acid sequence comprises at least two iterations of at least two distinct epitope-encoding nucleic acid sequences.
29. The composition of any one of claims 16-27, wherein the at least one antigen-encoding nucleic acid sequence comprises at least two iterations of at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, or at least 10 distinct epitope-encoding nucleic acid sequences.
30. The composition of any one of claims 16-29, wherein the at least two iterations are separated by at least one separate distinct epitope-encoding nucleic acid sequence.
31. The composition of any one of claims 16-29, wherein the at least two iterations are separated by at least 2 separate distinct epitope-encoding nucleic acid sequences.
32. The composition of any one of claims 16-29, wherein the at least two iterations, inclusive of the optional 5' linker sequence and/or the optional 3' linker sequence, are separated by at least 75 nucleotides.
33. The composition of any one of claims 16-29, wherein the at least two iterations, inclusive of the optional 5' linker sequence and/or the optional 3' linker sequence, are separated by at least 150 nucleotides, at least 300 nucleotides, or at least 675 nucleotides.

34. The composition of any one of claims 16-29, wherein the at least two iterations, inclusive of the optional 5' linker sequence and/or the optional 3' linker sequence, are separated by at least 50 nucleotides, at least 100 nucleotides, at least 200 nucleotides, at least 250 nucleotides, at least 350 nucleotides, at least 400 nucleotides, at least 450 nucleotides, at least 500 nucleotides, at least 700 nucleotides, at least 700 nucleotides, at least 750 nucleotides, at least 800 nucleotides, at least 900 nucleotides, or at least 1000 nucleotides.
35. The composition of any one of claims 16-29, wherein the at least two iterations, inclusive of the optional 5' linker sequence and/or the optional 3' linker sequence, are separated by at least 10 nucleotides, at least 15 nucleotides, at least 20 nucleotides, at least 25 nucleotides, at least 30 nucleotides, at least 35 nucleotides, at least 40 nucleotides, at least 45 nucleotides, at least 50 nucleotides, at least 55 nucleotides, at least 60 nucleotides, at least 65 nucleotides, or at least 70 nucleotides.
36. The composition of any one of claims 16-35, wherein the at least one antigen-encoding nucleic acid sequence is described, from 5' to 3', by the formula:
- $$(E_x-(E^N_n)_y)_z$$
- wherein,
- E represents a nucleotide sequence comprising at least one of the 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.
37. The composition of any one of claims 16-36, wherein the antigen-encoding cassette encodes at least 4 iterations of each of the amino acid sequences VVVGACGVGK (SEQ ID NO: 75), VVVGADGVGK (SEQ ID NO: 78), VVGAVGVGK (SEQ ID NO: 79), and ILDTAGHEEY (SEQ ID NO: 82).
38. The composition of any one of claims 16-36, wherein the distinct epitope-encoding nucleic acid sequences comprises at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, or at least 8 distinct epitope-encoding nucleic acid sequences each encoding distinct KRAS-associated MHC class I neoepitopes.

39. The composition of any one of claims 16-36, wherein each of the epitope-encoding nucleic acid sequences of the at least one antigen-encoding nucleic acid sequence encodes a distinct KRAS-associated MHC class I neoepitope.
40. The composition of any one of claims 16-39, wherein one or more of the nucleic acid sequences encoding the distinct KRAS-associated MHC class I neoepitopes comprises at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, or at least 8 iterations.
41. The composition of any one of claims 16-39, wherein each of the nucleic acid sequences encoding the distinct KRAS-associated MHC class I neoepitopes comprises at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, or at least 8 iterations.
42. The composition of any one of claims 16-39, wherein one or more of the nucleic acid sequences encoding the distinct KRAS-associated MHC class I neoepitopes comprises at least 4 iterations.
43. The composition of any one of claims 16-39, wherein each of the nucleic acid sequences encoding the distinct KRAS-associated MHC class I neoepitopes comprises at least 4 iterations.
44. The composition of any one of claims 16-43, wherein one or more of the distinct KRAS-associated MHC class I neoepitopes independently comprises a KRAS G12C mutation, a KRAS G12V mutation, a KRAS G12D mutation, or a KRAS Q61H mutation.
45. The composition of any one of claims 1-44, wherein the at least two iterations comprises a number of iterations, or z comprises a number, sufficient to stimulate a greater immune response relative to an antigen-encoding nucleic acid sequence comprising a single iteration of the epitope-encoding nucleic acid sequence.
46. The composition of any one of claims 1-45, wherein the at least two iterations comprises a number of iterations, or z comprises a number, sufficient to stimulate an immune response, and a single iteration of the epitope-encoding nucleic acid sequence is insufficient to stimulate the immune response or insufficient to stimulate a detectable immune response.
47. The composition of claims 45 or 46, wherein the immune response is an expansion of epitope-specific T cells following *in vivo* immunization with the composition for delivery of the antigen expression system.
48. The composition of claims 45 or 46, wherein the immune response is increased activation of epitope-specific T cells and/or increased epitope-specific killing by epitope-specific T cells following *in vivo* immunization with the composition for delivery of the antigen expression system.

49. A composition for delivery of an antigen expression system, comprising:
the antigen expression system,
wherein the antigen expression system comprises one or more vectors,
the one or more vectors comprising:
- (a) a vector backbone, wherein the backbone comprises:
 - (i) at least one promoter nucleotide sequence, and
 - (ii) optionally, at least one polyadenylation (poly(A)) sequence; and
 - (b) a cassette, wherein the cassette comprises:
 - (i) at least one antigen-encoding nucleic acid sequence, comprising:
 - (I) at least two distinct epitope-encoding nucleic acid sequences, 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, optionally wherein the at least one alteration is a KRAS mutation, 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, and wherein each of the epitope-encoding nucleic acid sequences comprises:
 - (A) optionally, a 5' linker sequence, and
 - (B) optionally, a 3' linker sequence;
 - (ii) optionally, a second promoter nucleotide sequence operably linked to the antigen-encoding nucleic acid sequence; and
 - (iii) optionally, at least one MHC class II epitope-encoding nucleic acid sequence;
 - (iv) optionally, at least one nucleic acid sequence encoding a GPGPG amino acid linker sequence (SEQ ID NO:56); and
 - (v) optionally, at least one second poly(A) sequence, wherein the second poly(A) sequence is a native poly(A) sequence or an exogenous poly(A) sequence to the vector backbone,

wherein if the second promoter nucleotide sequence is absent, the antigen-encoding nucleic acid sequence is operably linked to the at least one promoter nucleotide sequence, and

wherein the cassette does not encode an immunodominant MHC class I epitope that:

(1) stimulates a 5-fold or greater immune response when administered in a vaccine composition to a subject relative to another MHC class I epitope encoded in the cassette and capable of stimulating an immune response in the subject, and/or

(2) reduces an immune response to another MHC class I epitope encoded in the cassette when administered in a vaccine composition to a subject relative to an immune response when the other MHC class I epitope is administered in the absence of the immunodominant MHC class I epitope, optionally wherein the immune response is reduced to below a limit of detection and/or wherein the immune response is not a therapeutically effective response.

50. A composition for delivery of an antigen expression system, comprising:
the antigen expression system,

wherein the antigen expression system comprises one or more vectors,

the one or more vectors comprising:

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

(i) at least one promoter nucleotide sequence, and

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

(b) a cassette, wherein the cassette comprises:

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

(I) at least 2, 3, 4, 5, 6, 7, 8, 9, or 10 distinct epitope-encoding nucleic acid sequences linearly linked to each other,

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, optionally wherein the at least one alteration is a KRAS mutation, 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, and

wherein each of the epitope-encoding nucleic acid sequences comprises;

(A) optionally, a 5' linker sequence, and

(B) optionally, a 3' linker sequence;

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

(iii) optionally, at least one MHC class II epitope-encoding nucleic acid sequence;

(iv) optionally, at least one nucleic acid sequence encoding a GP GPG amino acid linker sequence (SEQ ID NO:56); and

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

wherein if the second promoter nucleotide sequence is absent, the antigen-encoding nucleic acid sequence is operably linked to the at least one promoter nucleotide sequence, and

wherein the cassette does not encode an immunodominant MHC class I epitope that:

(1) stimulates a 5-fold or greater immune response when administered in a vaccine composition to a subject relative to another MHC class I epitope encoded in the cassette and capable of stimulating an immune response in the subject, and/or

(2) reduces an immune response to another MHC class I epitope encoded in the cassette when administered in a vaccine composition to a subject relative to an immune response when the other MHC class I epitope is administered in the absence of the immunodominant MHC class I epitope, optionally wherein the immune response is reduced to below a limit of detection and/or wherein the immune response is not a therapeutically effective response.

51. The composition of claim 50, wherein at least one of the distinct epitope-encoding nucleic acid sequences encodes a KRAS-associated MHC class I neoepitope.

52. A composition for delivery of an antigen expression system, comprising:
the antigen expression system,

wherein the antigen expression system comprises one or more vectors,

the one or more vectors comprising:

(a) a vector backbone, wherein the vector backbone comprises a chimpanzee adenovirus vector, optionally wherein the chimpanzee adenovirus vector is a ChAdV68 vector, or an alphavirus vector, optionally wherein the alphavirus vector is a Venezuelan equine encephalitis virus vector; and

(b) a cassette, optionally wherein the cassette is integrated between a native promoter nucleotide sequence native to the vector backbone and a poly(A) sequence, optionally wherein the poly(A) sequence is native to the vector backbone, wherein the cassette comprises:

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

(I) at least one epitope-encoding nucleic acid sequence, optionally comprising at least two distinct epitope-encoding nucleic acid sequences linearly linked to each other, each epitope-encoding nucleic acid sequence optionally comprising:

(A) a MHC class I epitope encoding nucleic acid sequence, wherein the MHC class I epitope encoding nucleic acid sequence encodes a MHC class I epitope 7-15 amino acids in length,

(B) a 5' linker sequence, wherein the 5' linker sequence encodes a native N-terminal amino acid sequence of the MHC class I epitope, and wherein the 5' linker sequence encodes a peptide that is at least 2 amino acids in length,

(C) a 3' linker sequence, wherein the 3' linker sequence encodes a native C-terminal acid sequence of the MHC class I epitope, and wherein the 3' linker sequence encodes a peptide that is at least 2 amino acids in length, and

wherein the cassette is operably linked to the native promoter nucleotide sequence, wherein each of the epitope-encoding nucleic acid sequences encodes a polypeptide that is between 13 and 25 amino acids in length, and wherein each 3' end of each epitope-encoding nucleic acid sequence is linked to the 5' end of the following epitope-encoding nucleic acid sequence with the exception of the final epitope-encoding nucleic acid sequence in the cassette; and

(ii) at least two MHC class II epitope-encoding nucleic acid sequences comprising:

- (I) a PADRE MHC class II sequence (SEQ ID NO:48),
- (II) a Tetanus toxoid MHC class II sequence (SEQ ID NO:46),
- (III) a first nucleic acid sequence encoding a GPGPG amino acid linker sequence linking the PADRE MHC class II sequence and the Tetanus toxoid MHC class II sequence,
- (IV) a second nucleic acid sequence encoding a GPGPG amino acid linker sequence linking the 5' end of the at least two MHC class II epitope-encoding nucleic acid sequences to the epitope-encoding nucleic acid sequences,
- (V) optionally, a third nucleic acid sequence encoding a GPGPG amino acid linker sequence at the 3' end of the at least two MHC class II epitope-encoding nucleic acid sequences, and

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

wherein if the second promoter nucleotide sequence is absent, the antigen-encoding nucleic acid sequence is operably linked to the native promoter nucleotide sequence, and

wherein the cassette does not encode an immunodominant MHC class I epitope that:

(1) stimulates a 5-fold or greater immune response when administered in a vaccine composition to a subject relative to another MHC class I epitope encoded in the cassette and capable of stimulating an immune response in the subject, and/or

(2) reduces an immune response to another MHC class I epitope encoded in the cassette when administered in a vaccine composition to a subject relative to an immune response when the other MHC class I epitope is administered in the absence of the immunodominant MHC class I epitope, optionally wherein the immune response is reduced to below a limit of detection and/or wherein the immune response is not a therapeutically effective response.

53. The composition of any one of claims 49-52, wherein the immunodominant MHC class I epitope stimulates a 10-fold, 20-fold, 30-fold, 40-fold, 50-fold, 100-fold, 200-fold, 300-fold, 400-fold, 500-fold, 1000-fold, 2000-fold, 3000-fold, 4000-fold, 5000-fold, or 10,000-fold or greater immune response when administered in a vaccine composition to a subject relative to another MHC class I epitope encoded in the cassette and capable of stimulating an immune response in the subject.

54. The composition of any one of claims 49-52, wherein the immunodominant MHC class I epitope reduces the immune response of the other MHC class I epitope to below a limit of detection and/or does not stimulate a therapeutically effective response.
55. The composition of any one of claims 49-54, wherein the subject expresses at least one HLA allele known or predicted to present both the immunodominant MHC class I epitope and the other MHC class I epitope encoded in the cassette.
56. The composition of any one of claims 1-55, wherein one or more of the epitope-encoding nucleic acid sequences are derived from a tumor, an infection, or an infected cell of a subject.
57. The composition of any one of claims 1-55, wherein each of the epitope-encoding nucleic acid sequences are derived from a tumor, an infection, or an infected cell of a subject.
58. The composition of any one of claims 1-55, wherein one or more of the epitope-encoding nucleic acid sequences are not derived from a tumor, an infection, or an infected cell of a subject.
59. The composition of any one of claims 1-55, wherein each of the epitope-encoding nucleic acid sequences are not derived from a tumor, an infection, or an infected cell of a subject.
60. The composition of any one of claims 1-59, wherein the epitope-encoding nucleic acid sequence encodes an epitope known or suspected to be presented by MHC class I on a surface of a cell, optionally wherein the surface of the cell is a tumor cell surface or an infected cell surface, and optionally wherein the cell is a subject's cell.
61. The composition of claim 60, wherein the cell is a tumor cell 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, brain cancer, B-cell lymphoma, acute myelogenous leukemia, chronic myelogenous leukemia, chronic lymphocytic leukemia, T cell lymphocytic leukemia, non-small cell lung cancer, and small cell lung cancer, or wherein the cell is an infected cell selected from the group consisting of: a pathogen infected cell, a virally infected cell, a bacterially infected cell, a fungally infected cell, and a parasitically infected cell.
62. The composition of claim 61, wherein the virally infected cell is selected from the group consisting of: an HIV infected cell, a Severe acute respiratory syndrome-related coronavirus (SARS) infected cell, a severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infected cell, a Ebola infected cell, a Hepatitis B virus (HBV) infected cell, an influenza infected cell, an orthomyxoviridae family virus infected cell, a Human papillomavirus (HPV) infected cell, a Cytomegalovirus (CMV) infected cell, a

- Chikungunya virus infected cell, a Respiratory syncytial virus (RSV) infected cell, a Dengue virus infected cell, and a Hepatitis C virus (HCV) infected cell.
63. The composition of any one of the above claims, wherein the composition further comprises a nanoparticulate delivery vehicle.
64. The composition of claim 63, wherein the nanoparticulate delivery vehicle is a lipid nanoparticle (LNP).
65. The composition of claim 64, wherein the LNP comprises ionizable amino lipids.
66. The composition of claim 65, wherein the ionizable amino lipids comprise MC3-like (dilinoleylmethyl-4-dimethylaminobutyrate) molecules.
67. The composition of any of claims claim 24-66, wherein the nanoparticulate delivery vehicle encapsulates the antigen expression system.
68. The composition of any one of claims 16-18, 20-22, or 24-67, wherein the cassette is integrated between the at least one promoter nucleotide sequence and the at least one poly(A) sequence.
69. The composition of any one of claims 16-18, 20-22, or 24-68, wherein the second promoter is absent and the at least one promoter nucleotide sequence is operably linked to the antigen-encoding nucleic acid sequence.
70. The composition of any one of claims 16-18, 20-22, or 24-69, wherein the one or more vectors comprise one or more +-stranded RNA vectors.
71. The composition of claim 70 wherein the one or more +-stranded RNA vectors comprise a 5' 7-methylguanosine (m7g) cap.
72. The composition of claim 70 or 71, wherein the one or more +-stranded RNA vectors are produced by *in vitro* transcription.
73. The composition of any one of claims 16-18, 20-22, or 24-72, wherein the one or more vectors are self-replicating within a mammalian cell.
74. The composition of any one of claims 16-18, 20-22, or 24-73, wherein the backbone comprises at least one nucleotide sequence of an Aura virus, a Fort Morgan virus, a Venezuelan equine encephalitis virus, a Ross River virus, a Semliki Forest virus, a Sindbis virus, or a Mayaro virus.
75. The composition of any one of claims 16-18, 20-22, or 24-73, wherein the backbone comprises at least one nucleotide sequence of a Venezuelan equine encephalitis virus.
76. The composition of claim 74 or 75, wherein the backbone comprises at least sequences for nonstructural protein-mediated amplification, a 26S promoter sequence, a poly(A) sequence, a nonstructural protein 1 (nsP1) gene, a nsP2 gene, a nsP3 gene, and a nsP4 gene encoded by the nucleotide sequence of the Aura virus, the Fort Morgan virus, the

Venezuelan equine encephalitis virus, the Ross River virus, the Semliki Forest virus, the Sindbis virus, or the Mayaro virus.

77. The composition of claim 74 or 75, wherein the backbone comprises at least sequences for nonstructural protein-mediated amplification, a 26S promoter sequence, and a poly(A) sequence encoded by the nucleotide sequence of the Aura virus, the Fort Morgan virus, the Venezuelan equine encephalitis virus, the Ross River virus, the Semliki Forest virus, the Sindbis virus, or the Mayaro virus.
78. The composition of claim 76 or 77, 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.
79. The composition of any one of claims 76-78, wherein the backbone does not encode structural virion proteins capsid, E2 and E1.
80. The composition of claim 79, wherein the cassette is inserted in place of structural virion proteins within the nucleotide sequence of the Aura virus, the Fort Morgan virus, the Venezuelan equine encephalitis virus, the Ross River virus, the Semliki Forest virus, the Sindbis virus, or the Mayaro virus.
81. The composition of claim 74 or 75, wherein the Venezuelan equine encephalitis virus comprises the sequence of SEQ ID NO:3 or SEQ ID NO:5.
82. The composition of claim 74 or 75, 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.
83. The composition of claim 82, wherein the backbone comprises the sequence set forth in SEQ ID NO:6 or SEQ ID NO:7.
84. The composition of claim 82 or 83, wherein the cassette is inserted at position 7544 to replace the deletion between base pairs 7544 and 11175 as set forth in the sequence of SEQ ID NO:3 or SEQ ID NO:5.
85. The composition of claim 80-84, wherein the insertion of the cassette provides for transcription of a polycistronic RNA comprising the nsP1-4 genes and the at least one antigen-encoding nucleic acid sequence, wherein the nsP1-4 genes and the at least one antigen-encoding nucleic acid sequence are in separate open reading frames.
86. The composition of any one of claims 16-18, 20-22, or 24-73, wherein the backbone comprises at least one nucleotide sequence of a chimpanzee adenovirus vector.

87. The composition of claim 86, wherein the chimpanzee adenovirus vector is a ChAdV68 vector, optionally wherein the ChAdV68 vector comprises a ChAdV68 vector backbone comprising:
- the sequence set forth in SEQ ID NO:1;
 - the sequence set forth in SEQ ID NO:1, except that the sequence is fully deleted or functionally deleted in at least one gene selected from the group consisting of the chimpanzee adenovirus E1A, E1B, E2A, E2B, E3, E4, L1, L2, L3, L4, and L5 genes of the sequence set forth in SEQ ID NO:1, optionally wherein the sequence is fully deleted or functionally deleted in: (1) E1A and E1B; (2) E1A, E1B, and E3; or (3) E1A, E1B, E3, and E4 of the sequence set forth in SEQ ID NO:1;
 - a gene or regulatory sequence obtained from the sequence of SEQ ID NO:1, optionally wherein the gene is selected from the group consisting of the chimpanzee adenovirus inverted terminal repeat (ITR), E1A, E1B, E2A, E2B, E3, E4, L1, L2, L3, L4, and L5 genes of the sequence set forth in SEQ ID NO:1;
 - a partially deleted E4 gene comprising a deleted or partially-deleted E4orf2 region and a deleted or partially-deleted E4orf3 region, and optionally a deleted or partially-deleted E4orf4 region;
 - at least nucleotides 2 to 36,518 of the sequence set forth in SEQ ID NO:1 and further comprising: (1) an E1 deletion of at least nucleotides 577 to 3403 of the sequence shown in SEQ ID NO:1, (2) an E3 deletion of at least nucleotides 27,125 to 31,825 of the sequence shown in SEQ ID NO:1, and (3) an E4 deletion of at least nucleotides 34,916 to 35,642 of the sequence shown in SEQ ID NO:1; optionally wherein the antigen cassette is inserted within the E1 deletion;
 - the sequence set forth in SEQ ID NO:68, optionally wherein the antigen cassette is inserted within the E1 deletion;
 - one or more deletions between base pair number 577 and 3403 or between base pair 456 and 3014, and optionally wherein the vector further comprises one or more deletions between base pair 27,125 and 31,825 or between base pair 27,816 and 31,333 of the sequence set forth in SEQ ID NO:1; or
 - one or more deletions between base pair number 3957 and 10346, base pair number 21787 and 23370, and base pair number 33486 and 36193 of the sequence set forth in SEQ ID NO:1, and
- optionally wherein the cassette is inserted in the ChAdV vector backbone at the E1 region, E3 region, and/or any deleted AdV region that allows incorporation of the cassette.

88. The composition of any one of claims 16-18, 20-22, or 24-87, wherein the at least one promoter nucleotide sequence is the native 26S promoter nucleotide sequence encoded by the backbone.
89. The composition of any one of claims 16-18, 20-22, or 24-87, wherein the at least one promoter nucleotide sequence is an exogenous RNA promoter.
90. The composition of any one of claims 16-18, 20-22, or 24-89, wherein the second promoter nucleotide sequence is a 26S promoter nucleotide sequence.
91. The composition of any one of claims 16-18, 20-22, or 24-89, wherein the second promoter nucleotide sequence comprises multiple 26S promoter nucleotide sequences, wherein each 26S promoter nucleotide sequence provides for transcription of one or more of the separate open reading frames.
92. The composition of any one of the above claims, wherein the one or more vectors are each at least 300nt in size.
93. The composition of any one of the above claims, wherein the one or more vectors are each at least 1kb in size.
94. The composition of any one of the above claims, wherein the one or more vectors are each 2kb in size.
95. The composition of any one of the above claims, wherein the one or more vectors are each less than 5kb in size.
96. The composition of any one of the above claims, wherein at least one of the at least one antigen-encoding nucleic acid sequences encodes a polypeptide sequence or portion thereof that is presented by MHC class I on a cell surface, optionally a tumor cell surface or an infected cell surface.
97. The composition of any one of claims 1-18, 20-22, or 24-96, wherein each epitope-encoding nucleic acid sequence is linked directly to one another.
98. The composition of any one of claims 1-18, 20-22, or 24-97, wherein at least one of the epitope-encoding nucleic acid sequences is linked to a distinct epitope-encoding nucleic acid sequence with a nucleic acid sequence encoding a linker.
99. The composition of claim 98, wherein the linker links two MHC class I sequences or an MHC class I sequence to an MHC class II sequence.
100. The composition of claim 99, 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.
101. The composition of claim 98, wherein the linker links two MHC class II sequences or an MHC class II sequence to an MHC class I sequence.
 102. The composition of claim 101, wherein the linker comprises the sequence GPGPG.
 103. The composition of any one of claims 1-18, 20-22, or 24-102, wherein at least one sequence of the epitope-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 sequences of epitope encoded therefrom.
 104. The composition of claim 103, 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.
 105. The composition of any one of the above claims, wherein at least one of the epitope-encoding nucleic acid sequences encodes a polypeptide sequence or portion thereof that has increased binding affinity to its corresponding MHC allele relative to the translated, corresponding wild-type nucleic acid sequence.
 106. The composition of any one of the above claims, wherein at least one of the epitope-encoding nucleic acid sequences encodes a polypeptide sequence or portion thereof that has increased binding stability to its corresponding MHC allele relative to the translated, corresponding wild-type nucleic acid sequence.
 107. The composition of any one of the above claims, wherein at least one of the epitope-encoding nucleic acid sequences encodes a polypeptide sequence or portion thereof that has an increased likelihood of presentation on its corresponding MHC allele relative to the translated, corresponding wild-type nucleic acid sequence.
 108. The composition of any one of the above claims, wherein the at least one alteration comprises a point mutation, a frameshift mutation, a non-frameshift mutation, a deletion mutation, an insertion mutation, a splice variant, a genomic rearrangement, or a proteasome-generated spliced antigen.

109. The composition of any one of the above claims, wherein the tumor 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
or the infectious disease organism is selected from the group consisting of: Severe acute respiratory syndrome-related coronavirus (SARS), severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), Ebola, HIV, Hepatitis B virus (HBV), influenza, Hepatitis C virus (HCV), Human papillomavirus (HPV), Cytomegalovirus (CMV), Chikungunya virus, Respiratory syncytial virus (RSV), Dengue virus, an orthomyxoviridae family virus, and tuberculosis.
110. The composition of any one claims 1-18, 20-22, or 24-109, 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 epitope-encoding nucleic acid sequences.
111. The composition of any one of claims 1-18, 20-22, or 24-109, 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 epitope-encoding nucleic acid sequences.
112. The composition of any one of claims 1-18, 20-22, or 24-109, wherein the at least one antigen-encoding nucleic acid sequence comprises at least 2-400 epitope-encoding nucleic acid sequences and wherein at least two of the epitope-encoding nucleic acid sequences encode polypeptide sequences or portions thereof that are presented by MHC class I on a cell surface, optionally a tumor cell surface or an infected cell surface.
113. The composition of any one claims 1-18, 20-22, or 24-112, 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.
114. The composition of any one of claims 1-18, 20-22, or 24-112, 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.
115. The composition of any one of claims 1-18, 20-22, or 24-112, 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 polypeptide sequences or portions thereof that are presented by MHC class I on a cell surface, optionally a tumor cell surface or an infected cell surface.
116. The composition of claim 19 or 23, wherein at least two of the epitope-encoding nucleic acid sequences encode polypeptide sequences or portions thereof that are presented by MHC class I on a cell surface, optionally a tumor cell surface or an infected cell surface.
117. The composition of any one of the above claims, wherein when administered to the subject and translated, at least one of the epitopes encoded by the epitope-encoding nucleic acid sequence are presented on antigen presenting cells resulting in an immune response targeting at least one of the antigens on the tumor cell surface or the infected cell surface.
118. The composition of any one of the above claims, wherein the at least one antigen-encoding nucleic acid sequences when administered to the subject and translated, at least one of the MHC class I or class II epitopes are presented on antigen presenting cells resulting in an immune response targeting at least one of the epitopes on a tumor cell surface or the infected cell surface, and optionally wherein the expression of each of the at least one antigen-encoding nucleic acid sequences is driven by the at least one promoter nucleotide sequence.
119. The composition of any one of claims 1-18, 20-22, or 24-118, wherein each epitope-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.
120. The composition of any one of claims 1-18, 20-22, or 24-119, wherein the at least one MHC class II epitope-encoding nucleic acid sequence is present.
121. The composition of any one of claims 1-18, 20-22, or 24-119, 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 peptide sequence distinct from the corresponding peptide sequence encoded by a wild-type nucleic acid sequence.
122. The composition of any one of claims 1-18, 20-22, or 24-121, wherein the at least one MHC class II epitope-encoding nucleic acid sequence is 12-20, 12, 13, 14, 15, 16, 17, 18, 19, 20, or 20-40 amino acids in length.
123. The composition of any one of claims 1-18, 20-22, or 24-122, wherein the at least one MHC class II epitope-encoding nucleic acid sequence is present and comprises at least one universal MHC class II antigen-encoding nucleic acid sequence, optionally wherein the at least one universal sequence comprises at least one of Tetanus toxoid and PADRE.

124. The composition of any one of claims 16-18, 20-22, or 24-123, wherein the at least one promoter nucleotide sequence or the second promoter nucleotide sequence is inducible.
125. The composition of any one of claims 16-18, 20-22, or 24-123, wherein the at least one promoter nucleotide sequence or the second promoter nucleotide sequence is non-inducible.
126. The composition of any one of claims 16-18, 20-22, or 24-125, wherein the at least one poly(A) sequence comprises a poly(A) sequence native to the backbone.
127. The composition of any one of claims 16-18, 20-22, or 24-125, wherein the at least one poly(A) sequence comprises a poly(A) sequence exogenous to the backbone.
128. The composition of any one claims 16-18, 20-22, or 24-127, wherein the at least one poly(A) sequence is operably linked to at least one of the at least one antigen-encoding nucleic acid sequences.
129. The composition of any one of claims 16-18, 20-22, or 24-128, 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, or at least 100 consecutive A nucleotides.
130. The composition of any one of claims 16-18, 20-22, or 24-128, wherein the at least one poly(A) sequence is at least 80 consecutive A nucleotides.
131. The composition of any one of the above claims, wherein the cassette further comprises at least one of: an intron sequence, a woodchuck hepatitis virus posttranscriptional regulatory element (WPRE) sequence, an internal ribosome entry sequence (IRES) sequence, a nucleotide sequence encoding a 2A self-cleaving peptide sequence, a nucleotide sequence encoding a Furin cleavage site, or a sequence in the 5' or 3' non-coding region known to enhance the nuclear export, stability, or translation efficiency of mRNA that is operably linked to at least one of the at least one antigen-encoding nucleic acid sequences.
132. The composition of any one of the above claims, wherein the cassette further comprises a reporter gene, including but not limited to, green fluorescent protein (GFP), a GFP variant, secreted alkaline phosphatase, luciferase, a luciferase variant, or a detectable peptide or epitope.
133. The composition of claim 132, wherein the detectable peptide or epitope is selected from the group consisting of an HA tag, a Flag tag, a His-tag, or a V5 tag.
134. The composition of any one of the above claims, wherein the one or more vectors further comprises one or more nucleic acid sequences encoding at least one immune modulator.
135. The composition of claim 134, wherein the immune modulator is an anti-CTLA4 antibody or an antigen-binding fragment thereof, an anti-PD-1 antibody or an antigen-

- binding fragment thereof, an anti-PD-L1 antibody or an antigen-binding fragment thereof, an anti-4-1BB antibody or an antigen-binding fragment thereof, or an anti-OX-40 antibody or an antigen-binding fragment thereof.
136. The composition of claim 135, wherein the antibody or antigen-binding fragment thereof is a Fab fragment, a Fab' fragment, a single chain Fv (scFv), a single domain antibody (sdAb) either as single specific or multiple specificities linked together (e.g., camelid antibody domains), or full-length single-chain antibody (e.g., full-length IgG with heavy and light chains linked by a flexible linker).
137. The composition of claim 135 or 136, wherein the heavy and light chain sequences of the antibody are a contiguous sequence separated by either a self-cleaving sequence such as 2A or IRES; or the heavy and light chain sequences of the antibody are linked by a flexible linker such as consecutive glycine residues.
138. The composition of claim 134, wherein the immune modulator is a cytokine.
139. The composition of claim 138, wherein the cytokine is at least one of IL-2, IL-7, IL-12, IL-15, or IL-21 or variants thereof of each.
140. The composition of any one of claims 1-18, 20-22, or 24-139, wherein at least one 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 antigens;
 - (b) inputting the peptide sequence of each antigen into a presentation model to generate a set of numerical likelihoods that each of the antigens 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 antigens based on the set of numerical likelihoods to generate a set of selected antigens which are used to generate the epitope-encoding nucleic acid sequence.
141. The composition of claim 19 or 23, wherein each of the 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 antigens;

(b) inputting the peptide sequence of each antigen into a presentation model to generate a set of numerical likelihoods that each of the antigens 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 antigens based on the set of numerical likelihoods to generate a set of selected antigens which are used to generate the at least 20 epitope-encoding nucleic acid sequences.

142. The composition of claim 140, wherein a number of the set of selected antigens is 2-20.

143. The composition of claim 140-142, 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.

144. The composition of claim 140-143, wherein selecting the set of selected antigens comprises selecting antigens that have an increased likelihood of being presented on the cell surface relative to unselected antigens based on the presentation model, optionally wherein the selected antigens have been validated as being presented by one or more specific HLA alleles.

145. The composition of claim 140-144, wherein selecting the set of selected antigens comprises selecting antigens that have an increased likelihood of being capable of inducing a tumor-specific or infectious disease-specific immune response in the subject relative to unselected antigens based on the presentation model.

146. The composition of claim 140-145, wherein selecting the set of selected antigens comprises selecting antigens that have an increased likelihood of being capable of being presented to naïve T cells by professional antigen presenting cells (APCs) relative to unselected antigens based on the presentation model, optionally wherein the APC is a dendritic cell (DC).

147. The composition of claim 140-146, wherein selecting the set of selected antigens comprises selecting antigens that have a decreased likelihood of being subject to

- inhibition via central or peripheral tolerance relative to unselected antigens based on the presentation model.
148. The composition of claim 140-147, wherein selecting the set of selected antigens comprises selecting antigens that have a decreased likelihood of being capable of inducing an autoimmune response to normal tissue in the subject relative to unselected antigens based on the presentation model.
 149. The composition of claim 140-148, 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.
 150. The composition of claim 149, wherein the sequencing is next generation sequencing (NGS) or any massively parallel sequencing approach.
 151. The composition of any one of the above claims, wherein the cassette comprises junctional epitope sequences formed by adjacent sequences in the cassette.
 152. The composition of claim 151, wherein at least one or each junctional epitope sequence has an affinity of greater than 500 nM for MHC.
 153. The composition of claims 151 or 152, wherein each junctional epitope sequence is non-self.
 154. The composition of any one of the above claims, wherein each of the MHC class I epitopes is predicted or validated to be capable of presentation by at least one HLA allele present in at least 5% of a population.
 155. The composition of any one of the above claims, wherein each of the MHC class I epitopes is predicted or validated to be capable of presentation by at least one HLA allele, wherein each antigen/HLA pair has an antigen/HLA prevalence of at least 0.01% in a population.
 156. The composition of any one of the above claims, wherein each of the MHC class I epitopes is predicted or validated to be capable of presentation by at least one HLA allele, wherein each antigen/HLA pair has an antigen/HLA prevalence of at least 0.1% in a population.
 157. The composition of any one of the above claims, wherein the cassette does not encode a non-therapeutic MHC class I or class II epitope nucleic acid sequence comprising a translated, wild-type nucleic acid sequence, wherein the non-therapeutic epitope is predicted to be displayed on an MHC allele of the subject.
 158. The composition of claim 157, wherein the non-therapeutic predicted MHC class I or class II epitope sequence is a junctional epitope sequence formed by adjacent sequences in the cassette.

159. The composition of claims 151-158, wherein the prediction is based on presentation likelihoods generated by inputting sequences of the non-therapeutic epitopes into a presentation model.
160. The composition of any one of claims 151-159, wherein an order of the at least one antigen-encoding nucleic acid sequences in the cassette is determined by a series of steps comprising:
 - (a) generating a set of candidate cassette sequences corresponding to different orders of the at least one antigen-encoding nucleic acid sequences;
 - (b) determining, for each candidate cassette sequence, a presentation score based on presentation of non-therapeutic epitopes in the candidate cassette sequence; and
 - (c) selecting a candidate cassette sequence associated with a presentation score below a predetermined threshold as the cassette sequence for an antigen vaccine.
161. A pharmaceutical composition comprising the composition of any one of the above claims and a pharmaceutically acceptable carrier.
162. The composition of claim 161, wherein the composition further comprises an adjuvant.
163. The pharmaceutical composition of claim 161 or 162, wherein the composition further comprises an immune modulator.
164. The pharmaceutical composition of claim 163, wherein the immune modulator is an anti-CTLA4 antibody or an antigen-binding fragment thereof, an anti-PD-1 antibody or an antigen-binding fragment thereof, an anti-PD-L1 antibody or an antigen-binding fragment thereof, an anti-4-1BB antibody or an antigen-binding fragment thereof, or an anti-OX-40 antibody or an antigen-binding fragment thereof.
165. An isolated nucleotide sequence or set of isolated nucleotide sequences comprising the cassette of any of the above composition claims and one or more elements obtained from the sequence of SEQ ID NO:3 or SEQ ID NO:5, optionally wherein the one or more elements are selected from the group consisting of the sequences necessary for nonstructural protein-mediated amplification, the 26S promoter nucleotide sequence, the poly(A) sequence, and the nsP1-4 genes of the sequence set forth in SEQ ID NO:3 or SEQ ID NO:5, and optionally wherein the nucleotide sequence is cDNA.
166. The isolated nucleotide sequence of claim 165, wherein the sequence or set of isolated nucleotide sequences comprises the cassette of any of the above composition claims inserted at position 7544 of the sequence set forth in SEQ ID NO:6 or SEQ ID NO:7.
167. The isolated nucleotide sequence of claim 165 or 166, further comprising:
 - a) a T7 or SP6 RNA polymerase promoter nucleotide sequence 5' of the one or more elements obtained from the sequence of SEQ ID NO:3 or SEQ ID NO:5; and

- b) optionally, one or more restriction sites 3' of the poly(A) sequence.
168. The isolated nucleotide sequence of claim 165, wherein the cassette of any of the above composition claims is inserted at position 7563 of SEQ ID NO:8 or SEQ ID NO:9.
169. A vector or set of vectors comprising the nucleotide sequence of claims 165-168.
170. An isolated cell comprising the nucleotide sequence or set of isolated nucleotide sequences of claims 165-169, optionally wherein the cell is a BHK-21, CHO, HEK293 or variants thereof, 911, HeLa, A549, LP-293, PER.C6, or AE1-2a cell.
171. A kit comprising the composition of any of the above composition claims and instructions for use.
172. A method for treating a subject with cancer, the method comprising administering to the subject the composition of any of the above composition claims or the pharmaceutical composition of any of claims 161-164.
173. The method of claim 172, wherein the epitope-encoding nucleic acid sequence is derived from the tumor of the subject with cancer or from a cell or sample of the infected subject.
174. The method of claim 172, wherein the epitope-encoding nucleic acid sequence are not derived from the tumor of the subject with cancer or from a cell or sample of the infected subject.
175. A method for stimulating an immune response in a subject, the method comprising administering to the subject the composition of any of the above composition claims or the pharmaceutical composition of any of claims 161-164.
176. The method any of claims 172-175, wherein the subject expresses at least one HLA allele predicted or known to present the MHC class I epitope, optionally wherein the HLA is A*03:01, A*11:01, A*02:01, C*01:02, and/or A*01:01.
177. The method of any of claims 172-176, wherein the composition is administered intramuscularly (IM), intradermally (ID), subcutaneously (SC), or intravenously (IV).
178. The method of any of claims 172-176, wherein the composition is administered intramuscularly.
179. The method of any of claims 172-178, the method further comprising administration of one or more immune modulators, optionally wherein the immune modulator is administered before, concurrently with, or after administration of the composition or pharmaceutical composition.
180. The method of claim 179, wherein the one or more immune modulators are selected from the group consisting of: an anti-CTLA4 antibody or an antigen-binding fragment thereof, an anti-PD-1 antibody or an antigen-binding fragment thereof, an anti-PD-L1 antibody or

- an antigen-binding fragment thereof, an anti-4-1BB antibody or an antigen-binding fragment thereof, or an anti-OX-40 antibody or an antigen-binding fragment thereof.
181. The method of claim 179 or 180, wherein the immune modulator is administered intravenously (IV), intramuscularly (IM), intradermally (ID), or subcutaneously (SC).
182. The method of claim 181, wherein the subcutaneous administration is near the site of the composition or pharmaceutical composition administration or in close proximity to one or more vector or composition draining lymph nodes.
183. The method of any one of claims 172-182, further comprising administering to the subject a second vaccine composition.
184. The method of claim 183, wherein the second vaccine composition is administered prior to the administration of the composition or the pharmaceutical composition of any one of claims 172-182.
185. The method of claim 183, wherein the second vaccine composition is administered subsequent to the administration of the composition or the pharmaceutical composition of any one of claims 172-182.
186. The method of claim 184 or 185, wherein the second vaccine composition is the same as the composition or the pharmaceutical composition of any one of claims 172-182.
187. The method of claim 184 or 185, wherein the second vaccine composition is different from the composition or the pharmaceutical composition of any one of claims 172-182.
188. The method of claim 187, wherein the second vaccine composition comprises a chimpanzee adenovirus vector encoding at least one antigen-encoding nucleic acid sequence.
189. The method of claim 188, wherein the at least one antigen-encoding nucleic acid sequence encoded by the chimpanzee adenovirus vector is the same as the at least one antigen-encoding nucleic acid sequence of any of the above composition claims.
190. A method of manufacturing the one or more vectors of any of the above composition claims, the method comprising:
- (a) obtaining a linearized DNA sequence comprising the backbone and the cassette;
 - (b) *in vitro* transcribing the linearized DNA sequence by addition of the linearized DNA sequence to an *in vitro* transcription reaction containing all the necessary components to transcribe the linearized DNA sequence into RNA, optionally further comprising *in vitro* addition of the m7g cap to the resulting RNA; and
 - (c) isolating the one or more vectors from the *in vitro* transcription reaction.

191. The method of manufacturing of claim 190, wherein the linearized DNA sequence is generated by linearizing a DNA plasmid sequence or by amplification using PCR.
192. The method of manufacturing of claim 191, wherein the DNA plasmid sequence is generated using one of bacterial recombination or full genome DNA synthesis or full genome DNA synthesis with amplification of synthesized DNA in bacterial cells.
193. The method of manufacturing of claim 190, wherein isolating the one or more vectors from the *in vitro* transcription reaction involves one or more of phenol chloroform extraction, silica column based purification, or similar RNA purification methods.
194. A method of manufacturing the composition of any of the above composition claims for delivery of the antigen expression system, the method comprising:
 - (a) providing components for the nanoparticulate delivery vehicle;
 - (b) providing the antigen expression system; and
 - (c) providing conditions sufficient for the nanoparticulate delivery vehicle and the antigen expression system to produce the composition for delivery of the antigen expression system.
195. The method of manufacturing of claim 194, wherein the conditions are provided by microfluidic mixing.
196. A method for treating a subject with a disease, optionally wherein the disease is cancer or an infection, the method comprising administering to the subject an antigen-based vaccine to the subject, wherein the antigen-based vaccine comprises an antigen-encoding cassette, or a polypeptide sequence encoded by the cassette, wherein the antigen-encoding cassette comprises at least one antigen-encoding nucleic acid sequence described, from 5' to 3', by the formula:

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

wherein,

E represents a nucleotide sequence a 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, and

at least one of the distinct epitope-encoding nucleic acid sequences comprising the at least two iterations encodes a KRAS-associated MHC class I neoepitope.

197. A method for treating a subject with a disease, optionally wherein the disease is cancer, the method comprising administering to the subject an antigen-based vaccine to the subject, wherein the antigen-based vaccine comprises an antigen expression system, comprising:

the antigen expression system,

wherein the antigen expression system comprises one or more vectors,

the one or more vectors comprising:

- (a) a vector backbone, wherein the backbone comprises:

(i) at least one promoter nucleotide sequence, and

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

- (b) a cassette, wherein the cassette comprises:

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

(I) an epitope-encoding nucleic acid sequence encoding a KRAS-associated MHC class I neoepitope, and

wherein each of the epitope-encoding nucleic acid sequences comprises;

(A) optionally, a 5' linker sequence, and

(B) optionally, a 3' linker sequence;

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

- (iii) optionally, at least one MHC class II epitope-encoding nucleic acid sequence;

- (iv) optionally, at least one nucleic acid sequence encoding a GPGPG amino acid linker sequence (SEQ ID NO:56); and

- (v) optionally, at least one second poly(A) sequence, wherein the second poly(A) sequence is a native poly(A) sequence or an exogenous poly(A) sequence to the vector backbone,

wherein if the second promoter nucleotide sequence is absent, the antigen-encoding nucleic acid sequence is operably linked to the at least one promoter nucleotide sequence, and

wherein the at least one antigen-encoding nucleic acid sequence comprises at least two iterations of the epitope-encoding nucleic acid sequence encoding the KRAS-associated MHC class I neoepitope.

198. A method for treating a subject with a disease, optionally wherein the disease is cancer, the method comprising administering to the subject an antigen-based vaccine to the subject, wherein the antigen-based vaccine comprises an antigen expression system, comprising:
- the antigen expression system,
- wherein the antigen expression system comprises one or more vectors,
- the one or more vectors comprising:
- (a) a vector backbone, wherein the backbone comprises:
- (i) at least one promoter nucleotide sequence, and
- (ii) optionally, at least one polyadenylation (poly(A)) sequence; and
- (b) a cassette, wherein the cassette comprises:
- (i) at least one antigen-encoding nucleic acid sequence, comprising:
- (I) at least two distinct epitope-encoding nucleic acid sequences, 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, optionally wherein the at least one alteration is a KRAS mutation, 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, and wherein each of the epitope-encoding nucleic acid sequences comprises;
- (A) optionally, a 5' linker sequence, and
- (B) optionally, a 3' linker sequence;
- (ii) optionally, a second promoter nucleotide sequence operably linked to the antigen-encoding nucleic acid sequence; and
- (iii) optionally, at least one MHC class II epitope-encoding nucleic acid sequence;
- (iv) optionally, at least one nucleic acid sequence encoding a GPGPG amino acid linker sequence (SEQ ID NO:56); and

(v) optionally, at least one second poly(A) sequence, wherein the second poly(A) sequence is a native poly(A) sequence or an exogenous poly(A) sequence to the vector backbone,

wherein if the second promoter nucleotide sequence is absent, the antigen-encoding nucleic acid sequence is operably linked to the at least one promoter nucleotide sequence, and

wherein the cassette does not encode an immunodominant MHC class I epitope that:

(1) stimulates a 5-fold or greater immune response when administered in a vaccine composition to a subject relative to another MHC class I epitope encoded in the cassette and capable of stimulating an immune response in the subject, and/or

(2) reduces an immune response to another MHC class I epitope encoded in the cassette when administered in a vaccine composition to a subject relative to an immune response when the other MHC class I epitope is administered in the absence of the immunodominant MHC class I epitope, optionally wherein the immune response is reduced to below a limit of detection and/or wherein the immune response is not a therapeutically effective response.

199. The method of any one of claims 196-198, wherein the epitope-encoding nucleic acid sequence is derived from a tumor of the subject with cancer or from a cell or sample of the infected subject.
200. The method any one of claims 196-198, wherein the epitope-encoding nucleic acid sequence are not derived from a tumor of the subject with cancer or from a cell or sample of the infected subject.
201. A method for stimulating an immune response in a subject, the method comprising the method comprising administering to the subject an antigen-based vaccine to the subject, wherein the antigen-based vaccine comprises an antigen-encoding cassette, or a polypeptide sequence encoded by the cassette, wherein the antigen-encoding cassette comprises at least one antigen-encoding nucleic acid sequence described, from 5' to 3', by the formula:

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

wherein,

E represents a nucleotide sequence comprising a 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, and
at least one of the distinct epitope-encoding nucleic acid sequences comprising the at least two iterations encodes a KRAS-associated MHC class I neoepitope.
202. A method for stimulating an immune response in a subject, the method comprising the method comprising administering to the subject an antigen-based vaccine to the subject, wherein the antigen-based vaccine comprises:
an antigen expression system, comprising:
the antigen expression system,
wherein the antigen expression system comprises one or more vectors,
the one or more vectors comprising:
- (a) a vector backbone, wherein the backbone comprises:
 - (i) at least one promoter nucleotide sequence, and
 - (ii) optionally, at least one polyadenylation (poly(A)) sequence; and
 - (b) a cassette, wherein the cassette comprises:
 - (i) at least one antigen-encoding nucleic acid sequence, comprising:
 - (I) an epitope-encoding nucleic acid sequence encoding a KRAS-associated MHC class I neoepitope, and
wherein each of the epitope-encoding nucleic acid sequences comprises;
(A) optionally, a 5' linker sequence, and
(B) optionally, a 3' linker sequence;
 - (ii) optionally, a second promoter nucleotide sequence operably linked to the antigen-encoding nucleic acid sequence; and
 - (iii) optionally, at least one MHC class II epitope-encoding nucleic acid sequence;
 - (iv) optionally, at least one nucleic acid sequence encoding a GPGPG amino acid linker sequence (SEQ ID NO:56); and

(v) optionally, at least one second poly(A) sequence, wherein the second poly(A) sequence is a native poly(A) sequence or an exogenous poly(A) sequence to the vector backbone,

wherein if the second promoter nucleotide sequence is absent, the antigen-encoding nucleic acid sequence is operably linked to the at least one promoter nucleotide sequence, and

wherein the at least one antigen-encoding nucleic acid sequence comprises at least two iterations of at least one of the epitope-encoding nucleic acid sequences encoding the KRAS-associated MHC class I neoepitope.

203. A method for treating a subject with a disease, optionally wherein the disease is cancer of an infection, the method comprising administering to the subject an antigen-based vaccine to the subject, wherein the antigen-based vaccine comprises an antigen expression system, comprising:

the antigen expression system,

wherein the antigen expression system comprises one or more vectors,

the one or more vectors comprising:

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

(i) at least one promoter nucleotide sequence, and

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

(b) a cassette, wherein the cassette comprises:

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

(I) at least two distinct epitope-encoding nucleic acid sequences, 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, optionally wherein the at least one alteration is a KRAS mutation, 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, and wherein each of the epitope-encoding nucleic acid sequences comprises;

(A) optionally, a 5' linker sequence, and

(B) optionally, a 3' linker sequence;

- (ii) optionally, a second promoter nucleotide sequence operably linked to the antigen-encoding nucleic acid sequence; and
- (iii) optionally, at least one MHC class II epitope-encoding nucleic acid sequence;
- (iv) optionally, at least one nucleic acid sequence encoding a GP GPG amino acid linker sequence (SEQ ID NO:56); and
- (v) optionally, at least one second poly(A) sequence, wherein the second poly(A) sequence is a native poly(A) sequence or an exogenous poly(A) sequence to the vector backbone,

wherein if the second promoter nucleotide sequence is absent, the antigen-encoding nucleic acid sequence is operably linked to the at least one promoter nucleotide sequence, and

wherein the cassette does not encode an immunodominant MHC class I epitope that:

(1) stimulates a 5-fold or greater immune response when administered in a vaccine composition to a subject relative to another MHC class I epitope encoded in the cassette and capable of stimulating an immune response in the subject, and/or

(2) reduces an immune response to another MHC class I epitope encoded in the cassette when administered in a vaccine composition to a subject relative to an immune response when the other MHC class I epitope is administered in the absence of the immunodominant MHC class I epitope, optionally wherein the immune response is reduced to below a limit of detection and/or wherein the immune response is not a therapeutically effective response.

204. The method any of claims 196-203, wherein the subject expresses at least one HLA allele predicted or known to present the at least one epitope sequence, optionally wherein the at least one epitope sequence predicted or known to be presented comprises (1) the KRAS-associated MHC class I neoepitope, and/or (2) the immunodominant MHC class I epitope and the other MHC class I epitope encoded in the cassette.
205. The method any of claims 196-203, wherein the subject expresses at least one HLA allele predicted or known to present the at least one epitope sequence, and wherein the at least one epitope sequence comprises an epitope known or suspected to be presented by MHC class I on a surface of a cell, optionally wherein the at least one epitope sequence predicted or known to be presented comprises (1) the KRAS-associated MHC class I neoepitope, and/or (2) the immunodominant MHC class I epitope and the other MHC class I epitope encoded in the cassette.

206. The method of claim 205, wherein the surface of the cell is a tumor cell surface.
207. The method of claim 206, wherein the cell is a tumor cell 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, brain cancer, B-cell lymphoma, acute myelogenous leukemia, chronic myelogenous leukemia, chronic lymphocytic leukemia, T cell lymphocytic leukemia, non-small cell lung cancer, and small cell lung cancer.
208. The method of claim 205, wherein the surface of the cell is an infected cell surface.
209. The method of claim 208, wherein the cell is an infected cell selected from the group consisting of: a pathogen infected cell, a virally infected cell, a bacterially infected cell, a fungally infected cell, and a parasitically infected cell.
210. The method of claim 209, wherein the virally infected cell is selected from the group consisting of: an HIV infected cell, a Severe acute respiratory syndrome-related coronavirus (SARS) infected cell, a severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infected cell, a Ebola infected cell, a Hepatitis B virus (HBV) infected cell, an influenza infected cell, an orthomyxoviridae family virus infected cell, a Human papillomavirus (HPV) infected cell, a Cytomegalovirus (CMV) infected cell, a Chikungunya virus infected cell, a Respiratory syncytial virus (RSV) infected cell, a Dengue virus infected cell, and a Hepatitis C virus (HCV) infected cell.
211. A method for inducing an immune response in a subject, the method comprising administering to the subject an antigen-based vaccine to the subject, wherein the antigen-based vaccine comprises an antigen-encoding cassette, or a polypeptide sequence encoded by the cassette, wherein the antigen-encoding cassette comprises at least one antigen-encoding nucleic acid sequence described, from 5' to 3', by the formula:

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

wherein,

E represents a nucleotide sequence comprising a 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, and

at least one of the distinct epitope-encoding nucleic acid sequences comprising the at least two iterations encodes a KRAS-associated MHC class I neoepitope.

212. A method for inducing an immune response in a subject, the method comprising administering to the subject an antigen-based vaccine to the subject, wherein the antigen-based vaccine comprises:
- an antigen expression system, comprising:
 - the antigen expression system,
 - wherein the antigen expression system comprises one or more vectors,
 - the one or more vectors comprising:
 - (a) a vector backbone, wherein the backbone comprises:
 - (i) at least one promoter nucleotide sequence, and
 - (ii) optionally, at least one polyadenylation (poly(A)) sequence; and
 - (b) a cassette, wherein the cassette comprises:
 - (i) at least one antigen-encoding nucleic acid sequence, comprising:
 - (I) an epitope-encoding nucleic acid sequence encoding a KRAS-associated MHC class I neoepitope, and
 - wherein each of the epitope-encoding nucleic acid sequences comprises:
 - (A) optionally, a 5' linker sequence, and
 - (B) optionally, a 3' linker sequence;
 - (ii) optionally, a second promoter nucleotide sequence operably linked to the antigen-encoding nucleic acid sequence; and
 - (iii) optionally, at least one MHC class II epitope-encoding nucleic acid sequence;
 - (iv) optionally, at least one nucleic acid sequence encoding a GPGPG amino acid linker sequence (SEQ ID NO:56); and
 - (v) optionally, at least one second poly(A) sequence, wherein the second poly(A) sequence is a native poly(A) sequence or an exogenous poly(A) sequence to the vector backbone,
 - wherein if the second promoter nucleotide sequence is absent, the antigen-encoding nucleic acid sequence is operably linked to the at least one promoter nucleotide sequence,

wherein the at least one antigen-encoding nucleic acid sequence comprises at least two iterations of the epitope-encoding nucleic acid sequence encoding the KRAS-associated MHC class I neoepitope, and

wherein the subject expresses at least one HLA allele predicted or known to present the at least one KRAS-associated MHC class I neoepitope.

213. A method for inducing an immune response in a subject, the method comprising administering to the subject an antigen-based vaccine to the subject, wherein the antigen-based vaccine comprises:

an antigen expression system, comprising:

the antigen expression system,

wherein the antigen expression system comprises one or more vectors,

the one or more vectors comprising:

- (a) a vector backbone, wherein the backbone comprises:

(i) at least one promoter nucleotide sequence, and

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

- (b) a cassette, wherein the cassette comprises:

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

(I) at least two 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, optionally wherein the at least one alteration is a KRAS mutation, 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, and optionally wherein the epitope-encoding nucleic acid sequence encodes a MHC class I epitope, and

wherein each of the epitope-encoding nucleic acid sequences comprises;

(A) optionally, a 5' linker sequence, and

(B) optionally, a 3' linker sequence;

- (ii) optionally, a second promoter nucleotide sequence operably linked to the antigen-encoding nucleic acid sequence; and
- (iii) optionally, at least one MHC class II epitope-encoding nucleic acid sequence;
- (iv) optionally, at least one nucleic acid sequence encoding a GP GPG amino acid linker sequence (SEQ ID NO:56); and
- (v) optionally, at least one second poly(A) sequence, wherein the second poly(A) sequence is a native poly(A) sequence or an exogenous poly(A) sequence to the vector backbone,

wherein if the second promoter nucleotide sequence is absent, the antigen-encoding nucleic acid sequence is operably linked to the at least one promoter nucleotide sequence, and

wherein the cassette does not encode an immunodominant MHC class I epitope that:

(1) stimulates a 5-fold or greater immune response when administered in a vaccine composition to a subject relative to another MHC class I epitope encoded in the cassette and capable of stimulating an immune response in the subject, and/or

(2) reduces an immune response to another MHC class I epitope encoded in the cassette when administered in a vaccine composition to a subject relative to an immune response when the other MHC class I epitope is administered in the absence of the immunodominant MHC class I epitope, optionally wherein the immune response is reduced to below a limit of detection and/or wherein the immune response is not a therapeutically effective response, and

wherein the subject expresses at least one HLA allele predicted or known to present both the immunodominant MHC class I epitope and the other MHC class I epitope encoded in the cassette.

- 214. The method of any of claims 196-213, wherein the antigen expression system comprises any one of the antigen expression systems in any one of claims 1-160.
- 215. The method of any of claims 196-213, wherein the antigen-based vaccine comprises any one of the pharmaceutical compositions in any one of claims 161-164.
- 216. The method of any of claims 196-215, wherein the antigen-based vaccine is administered as a priming dose.
- 217. The method of any of claims 196-216, wherein the antigen-based vaccine is administered as one or more boosting doses.
- 218. The method of claim 217, wherein the boosting dose is different than the priming dose.

219. The method of claim 218, wherein:
- a) the priming dose comprises a chimpanzee adenovirus vector and the boosting dose comprises an alphavirus vector; or
 - b) the priming dose comprises an alphavirus vector and the boosting dose comprises a chimpanzee adenovirus vector.
220. The method of claim 217, wherein the boosting dose is the same as the priming dose.
221. The method of any one of claims 217-220, wherein the injection site of the one or more boosting doses is as close as possible to the injection site of the priming dose.
222. The method of any one of the above method claims, further comprising determining or having determined the HLA-haplotype of the subject.
223. The method of any one of the above method claims, wherein the antigen-based vaccine is administered intramuscularly (IM), intradermally (ID), subcutaneously (SC), or intravenously (IV).
224. The method of any one of the above method claims, wherein the antigen-based vaccine is administered intramuscularly (IM).
225. The method of claim 224, wherein the IM administration is administered at separate injection sites.
226. The method of claim 225, wherein the separate injection sites are in opposing deltoid muscles.
227. The method of claim 226, wherein the separate injection sites are in gluteus or rectus femoris sites on each side.
228. The method or composition of any of the above claims, wherein the KRAS-associated MHC class I neoepitope or the KRAS mutation comprises a KRAS G12C mutation, a KRAS G12V mutation, a KRAS G12D mutation, or a KRAS Q61H mutation.
229. The method or composition of any of the above claims, wherein the KRAS-associated MHC class I neoepitope or the KRAS mutation comprises any one of the amino acid sequence shown in SEQ ID NOs: 75-82.
230. The method or composition of any of the above claims, wherein the antigen-encoding cassette comprises each of the amino acid sequence shown in SEQ ID NOs: 75-82.
231. The method or composition of any of the above claims, wherein the antigen-encoding cassette comprises two or more iterations of each of the amino acid sequence shown in SEQ ID NOs: 75-82, optionally comprising 4 iterations of each of the amino acid sequence shown in SEQ ID NOs: 75-82.

232. The method or composition of any of the above claims, wherein the KRAS-associated MHC class I neoepitope or the KRAS mutation comprises the amino acid sequence shown in SEQ ID NO: 57, SEQ ID NO: 58, SEQ ID NO: 59, or SEQ ID NO: 60.
233. The method or composition of any of the above claims, wherein the epitope-encoding nucleic acid sequences comprises two or more distinct epitope-encoding nucleic acid sequences independently encoding a distinct KRAS-associated MHC class I neoepitope or a distinct KRAS mutation.
234. The method or composition of any of the above claims, wherein each of the epitope-encoding nucleic acid sequences independently encodes a distinct KRAS-associated MHC class I neoepitope or a distinct KRAS mutation.
235. The method or composition of any of the above claims, wherein the epitope-encoding nucleic acid sequences comprises two or more distinct epitope-encoding nucleic acid sequences independently encoding a KRAS G12C mutation, a KRAS G12V mutation, a KRAS G12D mutation, or a KRAS Q61H mutation.
236. The method or composition of any of the above claims, wherein the epitope-encoding nucleic acid sequences independently encodes each of a KRAS G12C mutation, a KRAS G12V mutation, and a KRAS G12D mutation, and optionally a KRAS Q61H mutation.
237. The method or composition of any of the above claims, wherein the antigen-encoding nucleic acid sequence encodes a peptide comprising the amino acid sequence shown in SEQ ID NO: 64 or SEQ ID NO: 65.
238. The method or composition of any of the above claims, wherein the cassette does not encode an immunodominant MHC class I epitope that:
- (1) stimulates a 5-fold or greater immune response when administered in a vaccine composition to a subject relative to another MHC class I epitope encoded in the cassette and capable of stimulating an immune response in the subject, and/or
 - (2) reduces an immune response to another MHC class I epitope encoded in the cassette when administered in a vaccine composition to a subject relative to an immune response when the other MHC class I epitope is administered in the absence of the immunodominant MHC class I epitope, optionally wherein the immune response is reduced to below a limit of detection and/or wherein the immune response is not a therapeutically effective response.
239. The method or composition of claim 236, wherein the cassette does not encode an immunodominant MHC class I epitope that stimulates a 5-fold or greater immune response when administered in a vaccine composition to a subject relative to a KRAS-

associated neoepitope encoded in the cassette and capable of stimulating an immune response in the subject.

240. The method or composition of claim 236, wherein the cassette does not encode an immunodominant MHC class I epitope that reduces an immune response to another MHC class I epitope encoded in the cassette when administered in a vaccine composition to a subject relative to an immune response when the other MHC class I epitope is administered in the absence of the immunodominant MHC class I epitope.
241. The method or composition of claim 236, wherein the cassette does not encode an immunodominant MHC class I epitope that reduces an immune response to another MHC class I epitope encoded in the cassette to below a limit of detection when administered in a vaccine composition to a subject relative to an immune response when the other MHC class I epitope is administered in the absence of the immunodominant MHC class I epitope.
242. The method or composition of claim 236, wherein the cassette does not encode an immunodominant MHC class I epitope that reduces an immune response to another MHC class I epitope encoded in the cassette when administered in a vaccine composition to a subject relative to an immune response when the other MHC class I epitope is administered in the absence of the immunodominant MHC class I epitope, wherein the immune response to the other MHC class I epitope is not a therapeutically effective response.
243. The method or composition of any of the above claims, wherein the immunodominant epitope is a TP53-associated MHC class I neoepitope, optionally wherein the TP53-associated MHC class I neoepitope comprises a S127Y mutation.

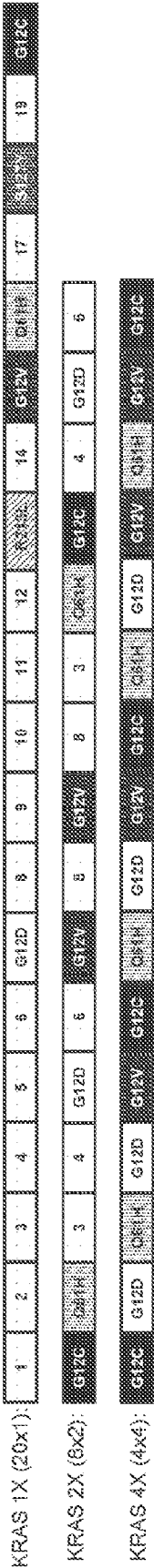


FIG. 1A

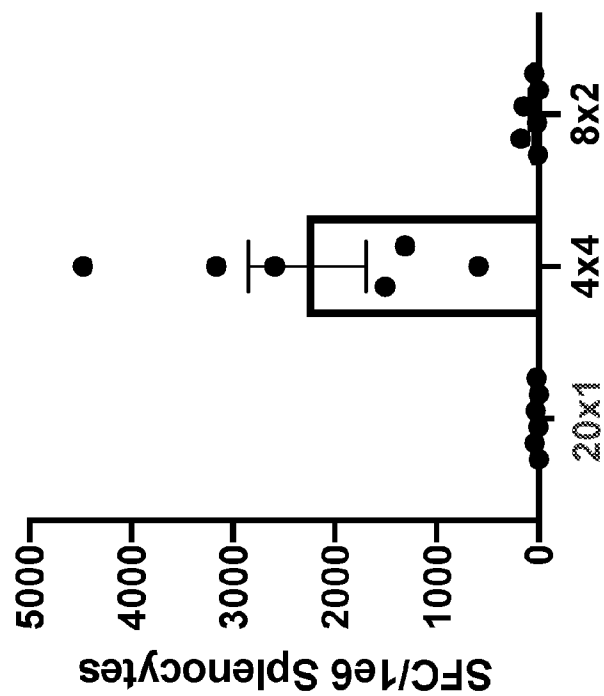


FIG. 1B

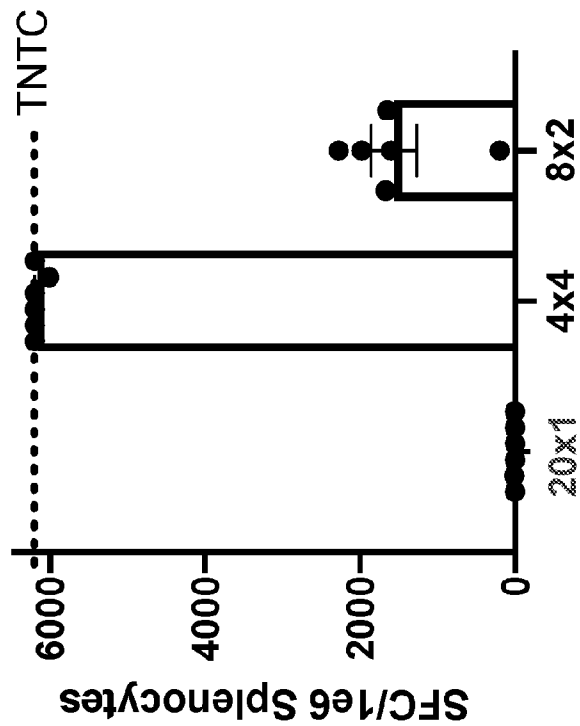


FIG. 1C

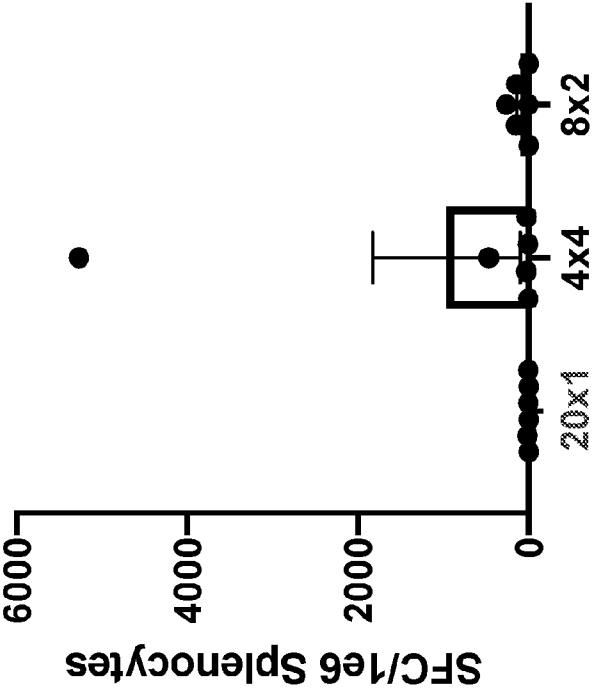


FIG. 1D

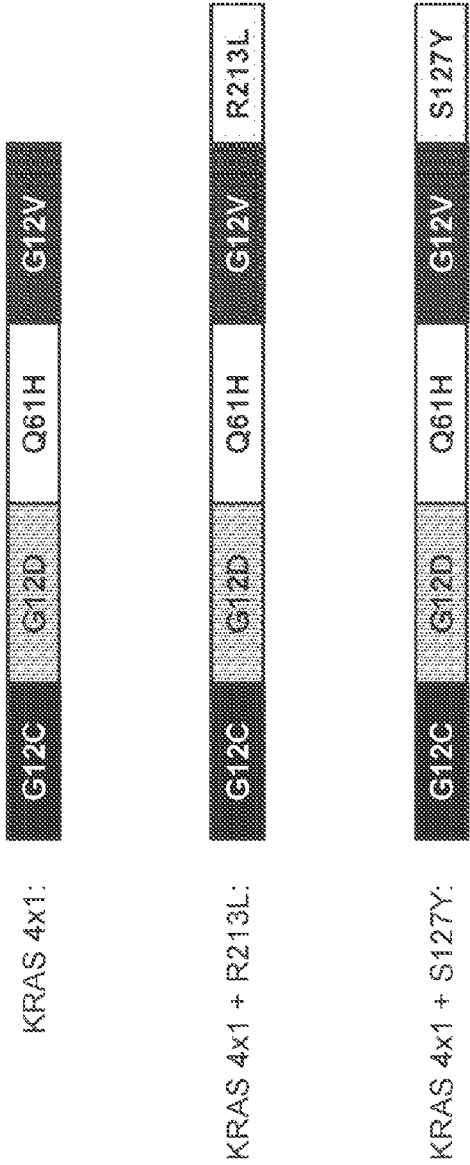


FIG. 2A

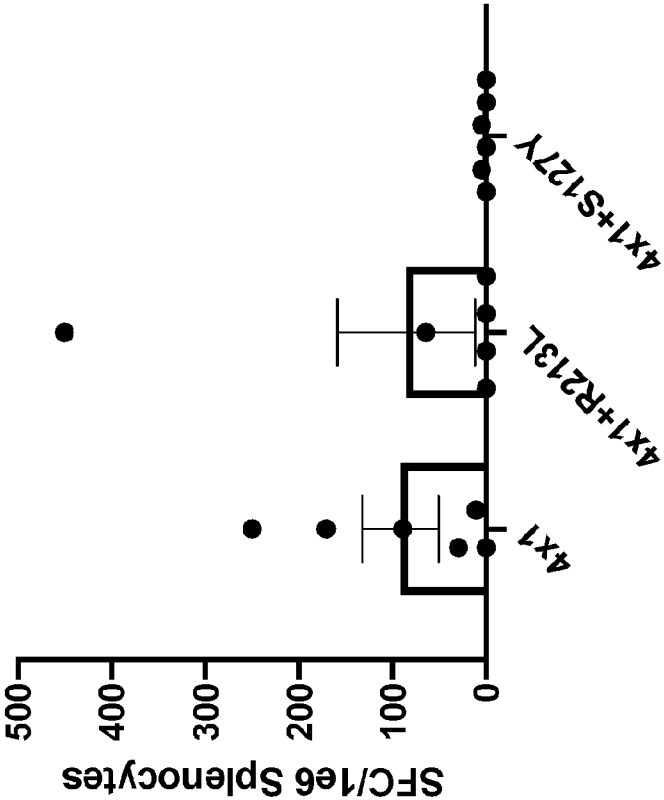
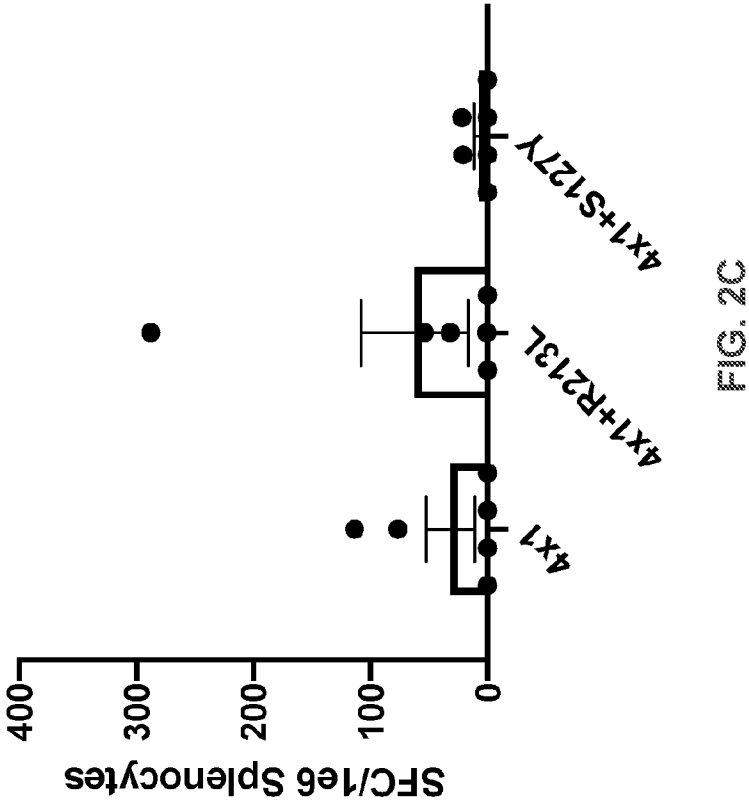


FIG. 2B



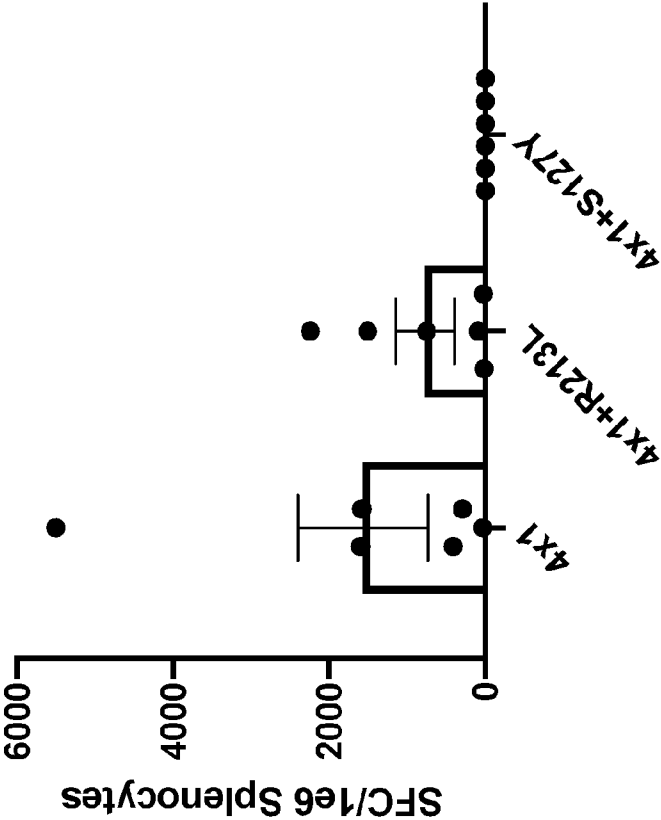
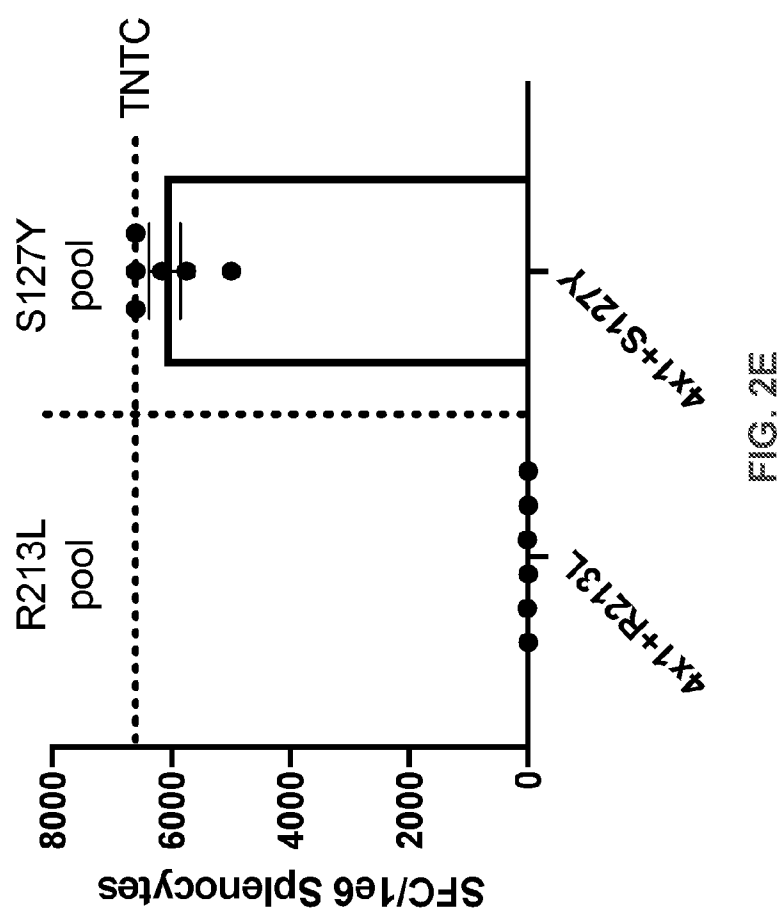
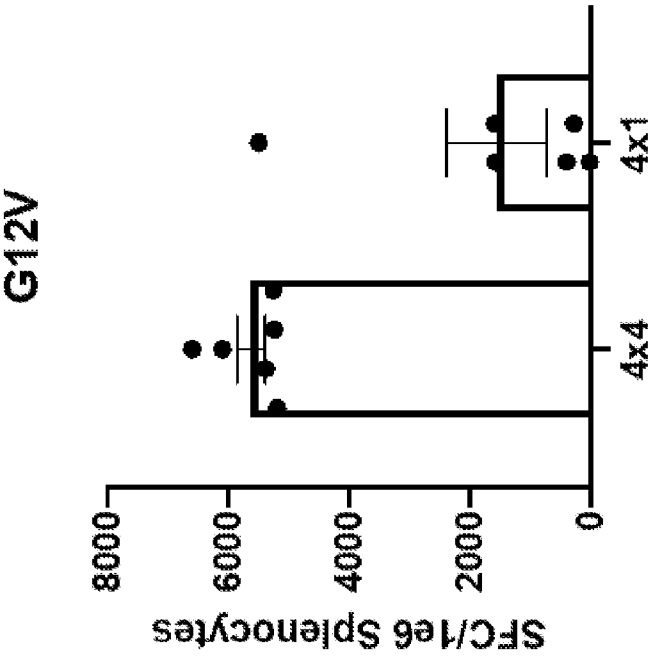
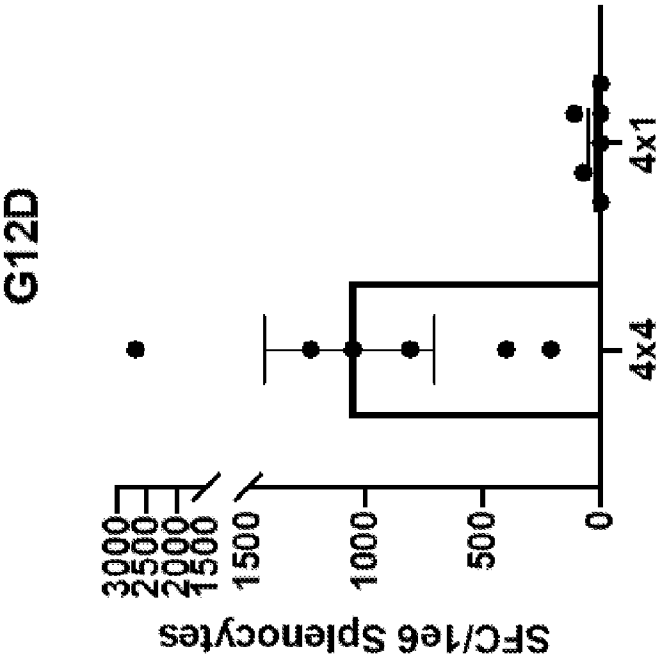


FIG. 2D





HLA-A*11 mice

FIG. 3

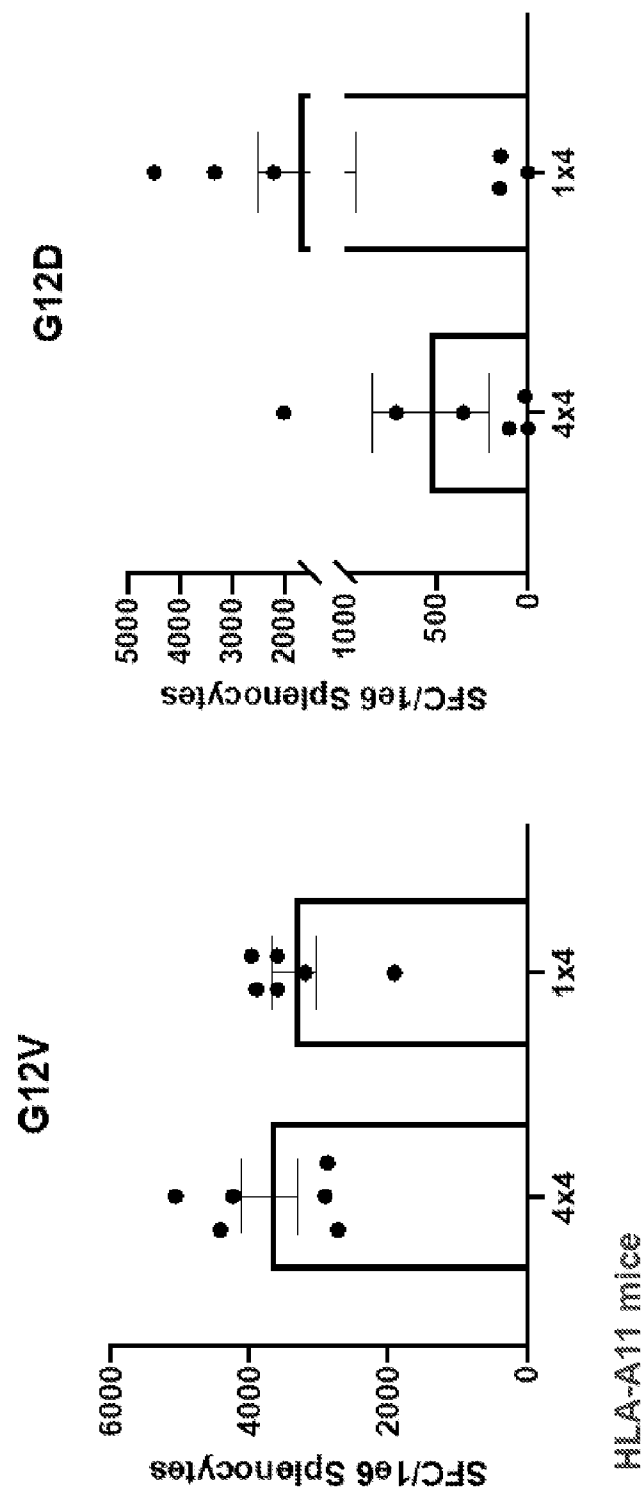


FIG. 4

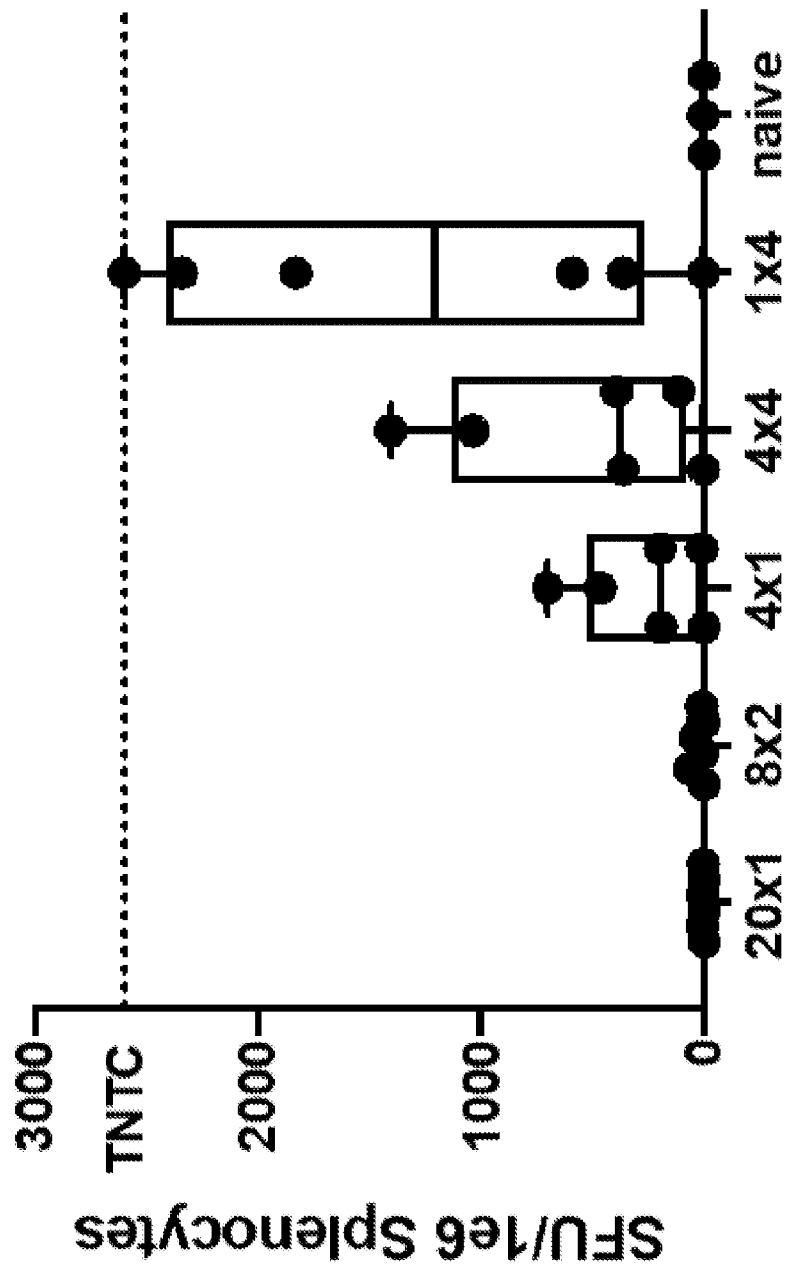


FIG. 5

HLA-A01 mice

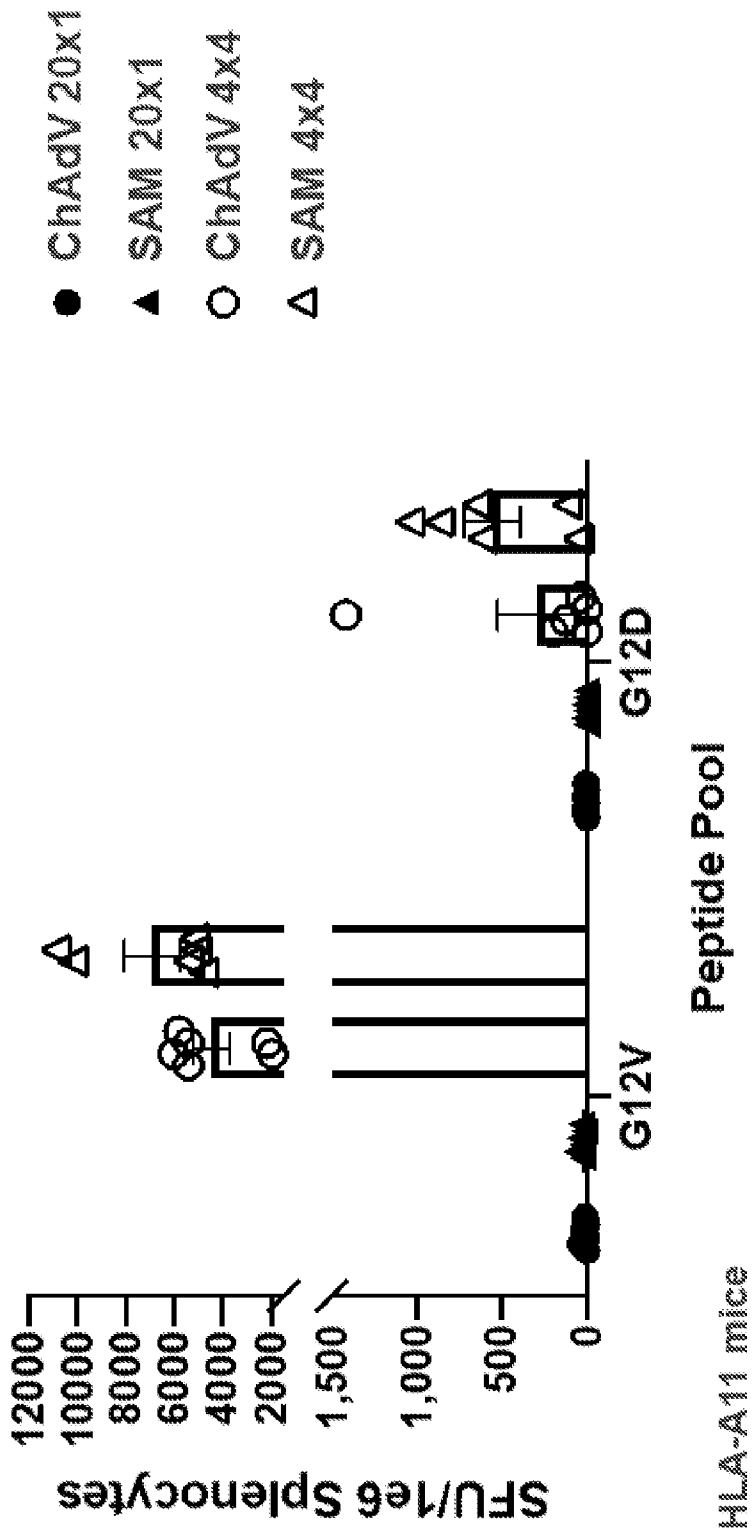
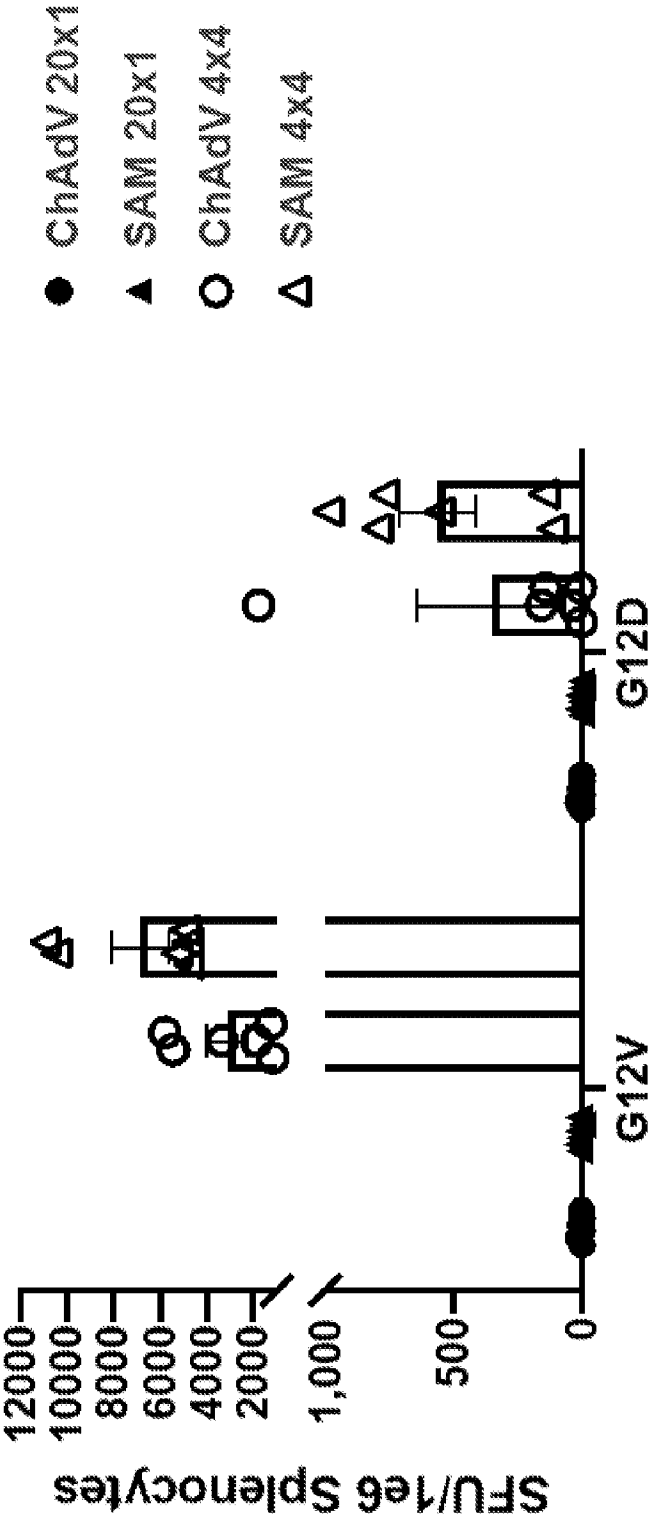


FIG. 6



HLA-A11 mice: Individual peptides

FIG. 7

SEQUENCE LISTING

<110> GRITSTONE ONCOLOGY, INC.

<120> MULTIEPITOPE VACCINE CASSETTES

<130> GSO-095

<140>

<141>

<150>

<151>

<160> 56

<170> PatentIn version 3.5

<210> 1

<211> 36519

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 1

ccatcttcaa taatatacct caaacttttt gtgcgcgtta atatgcaaat gaggcgtttg	60
aatttgggga ggaagggcgg tgattggtcg agggatgagc gaccgttagg ggcggggcga	120
gtgacgtttt gatgacgtgg ttgcgaggag gagccagttt gcaagttctc gtgggaaaag	180
tgacgtcaaa cgaggtgtgg tttgaacacg gaaatactca attttcccgc gctctctgac	240
aggaaatgag gtgtttctgg gcggatgcaa gtgaaaacgg gccattttcg cgcgaaaact	300
gaatgaggaa gtgaaaatct gagtaatttc gcgtttatgg caggaggagg tatttgccga	360
gggccgagta gactttgacc gattacgtgg gggtttcgat taccgtgttt ttcacctaaa	420
tttccgcgta cgggtgtcaaa gtccggtgtt tttacgtagg tgtcagctga tcgccagggt	480
atttaaacct gcgctctcca gtcaagaggc cactcttgag tgccagcgag aagagttttc	540
tcctccgcgc cgcgagtcag atctacactt tgaaagatga ggcacctgag agacctgccc	600
gatgagaaaa tcatcatcgc ttccgggaac gagattctgg aactggtggt aaatgccatg	660
atgggcgcag accctccgga gccccccacc ccatttgaga caccttcgct gcacgatttg	720
tatgatctgg aggtggatgt gcccgaggac gatcccaatg aggaggcggt aaatgatttt	780

ttagcgatg ccgcgctgct agctgccgag gaggcttcga gctctagctc agacagcgac	840
tcttctactgc atacccctag acccggcaga ggtgagaaaa agatccccga gcttaaaggg	900
gaagagatgg acttgcgctg ctatgaggaa tgcttgcccc cgagcgatga tgaggacgag	960
caggcgatcc agaacgcagc gagccaggga gtgcaagccg ccagcgagag ctttgcgctg	1020
gactgccccg ctctgcccgg acacggctgt aagtcttggtg aatttcatcg catgaatact	1080
ggagataaag ctgtgttggtg tgcactttgc tatatgagag cttacaacca ttgtgtttac	1140
agtaagtgtg attaagttga acttttagagg gaggcagaga gcagggtgac tgggcgatga	1200
ctggttttatt tatgtatata tgctctttat atagggtccc tctctgacgc agatgatgag	1260
acccccacta caaagtccac ttcgtcacc ccagaaattg gcacatctcc acctgagaat	1320
attgttagac cagttcctgt tagagccact gggaggagag cagctgtgga atgtttggat	1380
gacttgctac aggggtgggtg tgaacctttg gacttggtgta cccggaaacg cccagggcac	1440
taagtgccac acatgtgtgt ttacttgagg tgatgtcagt atttataggg tgtggagtgc	1500
aataaaaaat gtgttgactt taagtgcgtg gtttatgact caggggtggg gactgtgagt	1560
atataagcag gtgcagacct gtgtgggttag ctgagagcgg catggagatt tggacggtct	1620
tggaagactt tcacaagact agacagctgc tagagaacgc ctcgaacgga gtctcttacc	1680
tgtggagatt ctgcttcggt ggcgacctag ctaggctagt ctacagggcc aaacaggatt	1740
atagtgaaca atttgagggt attttgagag agtgttctgg tctttttgac gctcttaact	1800
tgggccatca gtctcacttt aaccagagga tttcgagagc ctttgatttt actactcctg	1860
gcagaaccac tgcagcagta gccttttttg cttttattct tgacaaatgg agtcaagaaa	1920
cccatttcag cagggttac cagctggatt tcttagcagt agctttgtgg agaactgga	1980
agtgccagcg cctgaatgca atctccggct acttgccggt acagccgcta gacactctga	2040
ggatcctgaa tctccaggag agtcccaggg cacgccaacg tcgccagcag cagcagcagg	2100
aggaggatca agaagagaac ccgagagccg gcctggaccc tccggcggag gaggaggagt	2160
agctgacctg tttcctgaac tgcgccgggt gctgactagg tcttcgagtg gtcgggagag	2220
ggggattaag cgggagaggc atgatgagac taatcacaga actgaactga ctgtgggtct	2280
gatgagtcgc aagcgcccag aaacagtgtg gtggcatgag gtgcagtcga ctggcacaga	2340
tgagggtgtc gtgatgcatg agaggttttc tctagaacaa gtcaagactt gttggttaga	2400

gcctgaggat gattgggagg tagccatcag gaattatgcc aagctggctc tgaggccaga	2460
caagaagtac aagattacta agctgataaa tatcagaaat gcctgctaca tctcagggaa	2520
tggggctgaa gtggagatct gtctccagga aagggtggct ttcagatgct gcatgatgaa	2580
tatgtacccg ggagtgggtg gcatggatgg ggttaccttt atgaacatga ggttcagggg	2640
agatgggtat aatggcacgg tctttatggc caataccaag ctgacagtcc atggctgctc	2700
cttctttggg ttttaataaca cctgcatcga ggcctggggg caggtcgggt tgaggggctg	2760
cagtttttca gccaaactgga tgggggtcgt gggcaggacc aagagtatgc tgtccgtgaa	2820
gaaatgcttg tttgagaggt gccacctggg ggtgatgagc gagggcgaag ccagaatccg	2880
ccactgcgcc tctaccgaga cgggctgctt tgtgctgtgc aagggaatg ctaagatcaa	2940
gcataatatg atctgtggag cctcggacga gcgcggctac cagatgctga cctgcgccgg	3000
cgggaacagc catatgctgg ccaccgtaca tgtggcttcc catgctcgca agccctggcc	3060
cgagttcgag cacaatgtca tgaccaggtg caatatgcat ctgggggtccc gccgaggcat	3120
gttcatgccc taccagtgc aacctgaatta tgtgaagggtg ctgctggagc ccgatgccat	3180
gtccagagtg agcctgacgg gggtgtttga catgaatgtg gaggtgtgga agattctgag	3240
atatgatgaa tccaagacca ggtgccgagc ctgcgagtgc ggagggaagc atgccaggtt	3300
ccagcccgtg tgtgtggatg tgacggagga cctgcgaccc gatcatttgg tgttgccctg	3360
caccgggacg gagttcgggt ccagcgggga agaactctgac tagagtgagt agtgttctgg	3420
ggcgggggag gacctgcatg agggccagaa taactgaaat ctgtgctttt ctgtgtgttg	3480
cagcagcatg agcggaagcg gctcctttga gggaggggta ttcagccctt atctgacggg	3540
gcgtctcccc tcctgggcgg gagtgcgtca gaatgtgatg ggatccacgg tggacggccg	3600
gcccgtgcag cccgcgaact cttcaaccct gacctatgca accctgagct cttcgtcggt	3660
ggacgcagct gccgccgag ctgctgcatc tgccgccagc gccgtgcgcg gaatggccat	3720
gggcgccggc tactacggca ctctggtggc caactcgagt tccaccaata atcccgccag	3780
cctgaacgag gagaagctgt tgctgctgat ggcccagctc gaggccttga cccagcgcct	3840
gggcgagctg acccagcagg tggctcagct gcaggagcag acgcgggccg cggttgccac	3900
ggtgaaatcc aaataaaaaa tgaatcaata aataaacgga gacggttggt gattttaaca	3960
cagagtctga atctttatctt gatctttcgc gcgcggtagg ccctggacca ccggtctcga	4020

tcattgagca	cccggtggat	cttttccagg	acccggtaga	ggtgggcttg	gatgttgagg	4080
tacatgggca	tgagcccgtc	ccgggggttg	aggtagctcc	attgcagggc	ctcgtgctcg	4140
ggggtggtgt	tgtaaatac	ccagtcata	caggggcgca	gggcatgggtg	ttgcacaata	4200
tctttgagga	ggagactgat	ggccacgggc	agcccttttg	tgtaggtgtt	tacaaatctg	4260
ttgagctggg	agggatgcat	gcggggggag	atgagggtgca	tcttggcctg	gatcttgaga	4320
ttggcgatgt	taccgcccag	atcccgctg	gggttcatgt	tgtgcaggac	caccagcacg	4380
gtgtatccgg	tgcacttggg	gaatttatca	tgcaacttgg	aagggaaggc	gtgaaagaat	4440
ttggcgacgc	ctttgtgccc	gccaggttt	tccatgcact	catccatgat	gatggcgatg	4500
ggcccgtggg	cggcggcctg	ggcaaagacg	tttcgggggt	cggacacatc	atagttgttg	4560
tcctgggtga	ggtcatcata	ggccatttta	atgaatttgg	ggcggagggt	gccggactgg	4620
gggacaaagg	taccctcgat	cccgggggcg	tagttcccct	cacagatctg	catctcccag	4680
gctttgagct	cggagggggg	gatcatgtcc	acctgcgggg	cgataaagaa	cacggtttcc	4740
ggggcggggg	agatgagctg	ggccgaaagc	aagtccgga	gcagctggga	cttgccgcag	4800
ccggtggggc	cgtagatgac	cccgatgacc	ggctgcagg	gtagttgag	ggagagacag	4860
ctgccgtcct	cccggaggag	gggggccacc	tcgttcatca	tctcgcgcac	gtgcatgttc	4920
tcgcgcacca	gttccgccag	gaggcgctct	ccccccagg	ataggagctc	ctggagcgag	4980
gcgaagtttt	tcagcggctt	gagtccgtcg	gccatgggca	ttttggagag	ggtttgttgc	5040
aagagtcca	ggcgggtcca	gagctcggtg	atgtgctcta	cggcatctcg	atccagcaga	5100
cctcctcgtt	tcgcgggttg	ggacggctgc	gggagtaggg	caccagacga	tgggcgtcca	5160
gcgcagccag	ggtccggtcc	ttccagggtc	gcagcgtccg	cgtcagggtg	gtctccgtca	5220
cggatgaagg	gtgcgcgccg	ggctgggcgc	ttgcgagggt	gcgcttcagg	ctcatccggc	5280
tggtcgaaaa	ccgctcccga	tcggcgccct	gcgcgtcggc	caggtagcaa	ttgacatga	5340
gttcgtagtt	gagcgctcg	gccgcgtggc	ctttggcgcg	gagcttacct	ttggaagtct	5400
gcccgcaggc	gggacagagg	agggacttga	gggcgtagag	cttgggggcg	aggaagacgg	5460
actcgggggc	gtaggcgtcc	gcgccgcagt	gggcgcagac	ggtctcgcac	tccacgagcc	5520
aggtgaggtc	gggctggtcg	gggtcaaaaa	ccagtttccc	gccgttcttt	ttgatgcgtt	5580
tcttaccttt	ggtctccatg	agctcgtgtc	cccgtgggt	gacaaagagg	ctgtccgtgt	5640

ccccgtagac cgactttatg ggccggtcct cgagcgggtgt gccgcgggtcc tcctcgtaga	5700
ggaaccccg c cactccgag acgaaagccc ggggtccaggc cagcacgaag gaggccacgt	5760
gggacgggta gcggtcgttg tccaccagcg ggtccacctt ttccagggtg tgcaaacaca	5820
tgtccccctc gtccacatcc aggaaggtga ttggcttgta agtgtaggcc acgtgaccgg	5880
gggtcccggc cgggggggta taaaagggtg cgggtccctg ctgctcctca ctgtcttccg	5940
gatcgtgtc caggagcgcc agctgttggg gtaggtattc cctctcgaag gcgggcatga	6000
cctcggcact caggttgtca gtttctagaa acgaggagga tttgatattg acggtgccgg	6060
cggagatgcc tttcaagagc ccctcgtcca tctggtcaga aaagacgatc tttttgttgt	6120
cgagcttggg ggcgaaggag ccgtagaggg cgttggagag gagcttggcg atggagcgca	6180
tggctctggtt tttttccttg tcggcgcgct ccttggcggc gatgttgagc tgcacgtact	6240
cgcgcgccac gcacttccat tcggggaaga cggtggtcag ctgctcgggc acgattctga	6300
cctgccagcc ccgattatgc agggatgatga ggtccacact ggtggccacc tcgccgcgca	6360
ggggctcatt agtccagcag aggcgtccgc ccttgcgca gcagaagggg ggcaaggggt	6420
ccagcatgac ctgctcgggg gggctggcat cgatggtgaa gatgccgggc aggaggtcgg	6480
ggtcaaagta gctgatggaa gtggccagat cgtccagggc agcttgccat tcgcgcacgg	6540
ccagcgcgcg ctgtaggga ctgaggggcg tgccccaggg catgggatgg gtaagcgcgg	6600
aggcgtacat gccgcagatg tcgtagacgt agaggggctc ctgaggatg ccgatgtagg	6660
tggggtagca gcgcccccg cggatgctgg cgcgcacgta gtcatacagc tcgtgcgagg	6720
gggcgaggag ccccgggccc aggttggtgc gactgggctt ttcggcgcg tagacgatct	6780
ggcggaaaat ggcatgcgag ttggaggaga tggtagggcct ttggaagatg ttgaagtggg	6840
cgtggggcag tccgaccgag tcgcggatga agtgggcgta ggagtcttgc agcttggcga	6900
cgagctcggc ggtgactagg acgtccagag cgcagtagtc gagggctctc tggatgatgt	6960
catacttgag ctgtcccttt tgtttccaca gctcgcggtt gagaaggaac tcttcgcggt	7020
ccttccagta ctcttcgagg gggaacccgt cctgatctgc acggtaagag cctagcatgt	7080
agaactgggt gacggccttg taggcgcagc agcccttctc cacggggagg gcgtaggcct	7140
gggcggcctt gcgcaggag gtgtgcgtga gggcgaaagt gtccctgacc atgacctga	7200
ggaactgggt cttgaagtcg atatcgtcgc agccccctg ctcccagagc tggaagtccg	7260

tgcgcttctt gtaggcgggg ttgggcaaag cgaaagtaac atcgttgaag aggatcttgc	7320
ccgcgcgggg cataaagttg cgagtgatgc ggaaaggttg gggcacctcg gcccggttgt	7380
tgatgacctg ggcggcgagc acgatctcgt cgaagccgtt gatgttggtg cccacgatgt	7440
agagttccac gaatcgcgga cggcccttga cgtggggcag tttcttgagc tcctcgtagg	7500
tgagctcgtc ggggtcgctg agcccgtgct gctcgagcgc ccagtcggcg agatgggggt	7560
tggcgcggag gaaggaagtc cagagatcca cggccagggc ggtttgaga cggtcccgg	7620
actgacggaa ctgctgccc acggccattt tttcgggggt gacgcagtag aaggcgcggg	7680
gggtcccgtg ccagcgatcc ctttgagct ggagggcgag atcgagggcg agctcgacga	7740
gccggctcgc cccggagagt ttcattgacca gcatgaagg gacgagctgc ttgccgaagg	7800
accccatcca ggtgtaggtt tccacatcgt aggtgaggaa gaggctttcg gtgcgaggat	7860
gcgagccgat ggggaagaac tggatctcct gccaccaatt ggaggaatgg ctgttgatgt	7920
gatggaagta gaaatgccga cggcgcgccg aacactcgtg cttgtgttta tacaagcggc	7980
cacagtgtc gcaacgtgc acgggatgca cgtgtgcac gagctgtacc tgagttcctt	8040
tgacgaggaa tttcagtggg aagtggagtc gtggcgctg catctcgtgc tgtactacgt	8100
cgtggtgtc ggcctggccc tcttctgcct cgatggtggt catgctgacg agccgcgcg	8160
ggaggcaggt ccagacctcg gcgcgagcgg gtcggagagc gaggacgagg gcgcgcaggc	8220
cggagctgtc cagggtcctg agacgtgcg gagtcaggtc agtgggcagc ggccgcgcgc	8280
ggttgacttg caggagtttt tccagggcgc gcgggaggtc cagatggtac ttgatctcca	8340
ccgcgccatt ggtggcgacg tcgatggctt gcagggtccc gtgcccctgg ggtgtgacca	8400
ccgtcccccg tttcttcttg ggcggctggg gcgacggggg cggcgcctct tccatggtta	8460
gaagcggcgg cgaggacgcg gccggggcgg caggggcggc tcggggcccg gaggcagggg	8520
cggcaggggc acgtcggcgc gcgcgcggg taggttcttg tactgcgcc ggagaagact	8580
ggcgtgagcg acgacgcgac ggttgacgtc ctggatctga cgcctctggg tgaaggccac	8640
gggacccgtg agtttgaacc tgaaagagag ttcgacagaa tcaatctcgg tatcgttgac	8700
ggcggcctgc cgcaggatct cttgcacgtc gcccgagttg tcctggtagg cgatctcgg	8760
catgaactgc tcgatctcct cctcttgaag gtctccgcgg ccggcgcgct ccacggtggc	8820
cgcgaggtcg ttggagatgc ggcccatgag ctgcgagaag gcgttcatgc ccgcctcgtt	8880

ccagacgcgg	ctgtagacca	cgacgccctc	gggatcgcgg	gcgcgcatga	ccacctgggc	8940
gaggttgagc	tccacgtggc	gcgtgaagac	cgcgtagttg	cagaggcgct	ggtagaggta	9000
gttgagcgtg	gtggcgatgt	gctcggtgac	gaagaaatac	atgatccagc	ggcggagcgg	9060
catctcgctg	acgtcgccca	gcgcctccaa	acgttccatg	gcctcgtaaa	agtccacggc	9120
gaagttgaaa	aactgggagt	tgcgcgccga	gacggtaaac	tcctcctcca	gaagacggat	9180
gagctcggcg	atggtggcgc	gcacctcgcg	ctcgaaggcc	cccgggagtt	cctccacttc	9240
ctcttcttcc	tcctccacta	acatctcttc	tacttctctc	tcaggcggca	gtggtggcgg	9300
gggagggggc	ctgcgtcgcc	ggcggcgcac	gggcagacgg	tcgatgaagc	gctcgatggt	9360
ctcgccgcgc	cggcgtcgca	tggctctcgg	gacggcgcgc	ccgtcctcgc	ggggccgcag	9420
cgtgaagacg	ccgccgcgca	tctccagggtg	gccggggggg	tccccgttgg	gcagggagag	9480
ggcgtgacg	atgcatctta	tcaattgccc	cgtagggact	ccgcgcaagg	acctgagcgt	9540
ctcgagatcc	acgggatctg	aaaaccgctg	aacgaaggct	tcgagccagt	cgcagtcgca	9600
aggtaggctg	agcacggttt	cttctggcgg	gtcatgttgg	ttgggagcgg	ggcgggagat	9660
gctgctggtg	atgaagttga	aataggcggt	tctgagacgg	cggatggtgg	cgaggagcac	9720
caggtctttg	ggcccggctt	gctggatgcg	cagacggtcg	gccatgcccc	aggcgtggtc	9780
ctgacacctg	gccaggctct	tgtagtagtc	ctgcatgagc	cgctccacgg	gcacctctc	9840
ctcgcccgcg	cggccgtgca	tgcgcgtgag	cccgaagccg	cgctggggct	ggacgagcgc	9900
caggtcggcg	acgacgcgct	cggcgaggat	ggcttgctgg	atctgggtga	gggtggtctg	9960
gaagtcatca	aagtgcacga	agcgggtgta	ggctccggtg	ttgatggtgt	aggagcagtt	10020
ggccatgacg	gaccagttga	cggctctggtg	gcccggacgc	acgagctcgt	ggtacttgag	10080
gcgcgagtag	gcgcgcgtgt	cgaagatgta	gtcgttgacg	gtgcgcacca	ggtactggtg	10140
gccgatgagg	aagtgcggcg	gcggctggcg	gtagagcggc	catcgctcgg	tggcgggggc	10200
gccgggcgcg	aggtcctcga	gcatggtgcg	gtggtagccg	tagatgtacc	tggacatcca	10260
ggtgatgccg	gcggcggtgg	tggaggcgcg	cgggaactcg	cggacgcggt	tccagatgtt	10320
gcgcagcggc	aggaagtagt	tcatggtggg	cacggtctgg	cccgtgaggc	gcgcgcagtc	10380
gtggatgctc	tatacgggca	aaaacgaaag	cggtcagcgg	ctcgactccg	tggcctggag	10440
gctaagcgaa	cgggttgggc	tgcgcgtgta	ccccggttcg	aatctcgaat	caggctggag	10500

ccgcagctaa	cgtggtattg	gcactcccgt	ctcgacccaa	gcctgcacca	accctccagg	10560
atacggaggc	gggtcgtttt	gcaacttttt	tttggaggcc	ggatgagact	agtaagcgcg	10620
gaaagcggcc	gaccgcgatg	gctcgctgcc	gtagtctgga	gaagaatcgc	cagggttgcg	10680
ttgcggtgtg	ccccggttcg	aggccggccg	gattccgcgg	ctaacgaggg	cgtggctgcc	10740
ccgtcgtttc	caagacccca	tagccagccg	acttctccag	ttacggagcg	agcccccttt	10800
ttgttttgtt	tgtttttgcc	agatgcatcc	cgtactgcgg	cagatgcgcc	cccaccaccc	10860
tccaccgcaa	caacagcccc	ctccacagcc	ggcgcttctg	cccccgcccc	agcagcaact	10920
tccagccacg	accgccgcgg	ccgccgtgag	cggggctgga	cagagttatg	atcaccagct	10980
ggccttgga	gagggcgagg	ggctggcgcg	cctggggggcg	tcgtcgccgg	agcggcaccc	11040
gcgcgtgcag	atgaaaaggg	acgctcgcga	ggcctacgtg	cccaagcaga	acctgttcag	11100
agacaggagc	ggcgaggagc	ccgaggagat	gcgcgcggcc	cggttccacg	cggggcgggga	11160
gctgcggcgc	ggcctggacc	gaaagagggt	gctgaggggac	gaggatttcg	aggcggacga	11220
gctgacgggg	atcagccccg	cgcgcgcgca	cgtggccgcg	gccaacctgg	tcacggcgta	11280
cgagcagacc	gtgaaggagg	agagcaactt	caaaaaatcc	ttcaacaacc	acgtgcgcac	11340
cctgatcgcg	cgcgaggagg	tgaccctggg	cctgatgcac	ctgtgggacc	tgctggaggc	11400
catcgtgcag	aacccccacca	gcaagccgct	gacggcgcgag	ctgttcctgg	tggtgcagca	11460
tagtcgggac	aacgaagcgt	tcagggaggc	gctgctgaat	atcaccgagc	ccgaggggccg	11520
ctggctcctg	gacctggtga	acattctgca	gagcatcgtg	gtgcaggagc	gcgggctgcc	11580
gctgtccgag	aagctggcgg	ccatcaactt	ctcggtgctg	agtttgggca	agtactacgc	11640
taggaagatc	tacaagaccc	cgtacgtgcc	catagacaag	gaggtgaaga	tcgacggggtt	11700
ttacatgcgc	atgaccctga	aagtgctgac	cctgagcgac	gatctggggg	tgtaccgcaa	11760
cgacaggatg	caccgtgcgg	tgagcgccag	caggcggcgc	gagctgagcg	accaggagct	11820
gatgcatagt	ctgcagcggg	ccctgaccgg	ggccggggacc	gagggggaga	gctactttga	11880
catgggcgcg	gacctgcact	ggcagcccag	ccgccggggcc	ttggaggcgg	cggcaggacc	11940
ctacgtagaa	gaggtggacg	atgaggtgga	cgaggagggc	gagtacctgg	aagactgatg	12000
gcgcgaccgt	atTTTTgcta	gatgcaacaa	caacagccac	ctcctgatcc	cgcgatgcgg	12060
gcggcgctgc	agagccagcc	gtccggcatt	aactcctcgg	acgattggac	ccaggccatg	12120

caacgcatca	tggcgctgac	gacccgcaac	cccgaagcct	ttagacagca	gccccaggcc	12180
aaccggctct	cggccatcct	ggaggccgtg	gtgccctcgc	gctccaaccc	cacgcacgag	12240
aaggtcctgg	ccatcgtgaa	cgcgctggtg	gagaacaagg	ccatccgcgg	cgacgaggcc	12300
ggcctggtgt	acaacgcgct	gctggagcgc	gtggcccgcct	acaacagcac	caacgtgcag	12360
accaacctgg	accgcatggt	gaccgacgtg	cgcgaggccg	tggcccagcg	cgagcggttc	12420
caccgcgagt	ccaacctggg	atccatggtg	gcgctgaacg	ccttcctcag	caccagccc	12480
gccaacgtgc	cccggggcca	ggaggactac	accaacttca	tcagcgccct	gcgctgatg	12540
gtgaccgagg	tgccccagag	cgaggtgtac	cagtccgggc	cggactactt	cttcagacc	12600
agtcgccagg	gcttgacagc	cgtgaacctg	agccaggctt	tcaagaactt	gcagggcctg	12660
tggggcgtgc	aggccccggt	cggggaccgc	gcgacggtgt	cgagcctgct	gacgccgaac	12720
tcg'gcctgc	tgctgctgct	ggtggcccc	ttcacggaca	gcggcagcat	caaccgcaac	12780
tcgtacctgg	gctacctgat	taacctgtac	cgcgaggcca	tcggccaggc	gcacgtggac	12840
gagcagacct	accaggagat	caccacgtg	agccgcgccc	tgggccagga	cgacccgggc	12900
aacctggaag	ccaccctgaa	ctttttgctg	accaaccggt	cgcagaagat	ccgccccag	12960
tacgcgctca	gcaccgagga	ggagcgcata	ctgcgttacg	tgcagcagag	cgtgggcctg	13020
ttcctgatgc	aggagggggc	cacccccagc	gccgcgctcg	acatgaccgc	gcgcaacatg	13080
gagcccagca	tgtacgccag	caaccgccc	ttcatcaata	aactgatgga	ctacttgcat	13140
cgggcggccg	ccatgaactc	tgactatttc	accaacgcca	tcctgaatcc	ccactggctc	13200
ccgccgccgg	ggttctacac	gggcgagtac	gacatgccc	accccaatga	cgggttcctg	13260
tgggacgatg	tggacagcag	cgtgttctcc	ccccgaccgg	gtgctaacga	gcgccccttg	13320
tggaagaagg	aaggcagcga	ccgacgccc	tcctcggcgc	tgtccggccg	cgagggtgct	13380
gccgcggcgg	tgcccagggc	cgccagtcct	ttcccagact	tgcccttctc	gctgaacagt	13440
atccgcagca	gcgagctggg	caggatcacg	cgccgcgct	tgctgggcga	agaggagtac	13500
ttgaatgact	cgctgttgag	acccgagcgg	gagaagaact	tccccaataa	cgggatagaa	13560
agcctggtgg	acaagatgag	ccgctggaag	acgtatgcgc	aggagcacag	ggacgatccc	13620
cgggcgtcgc	agggggccac	gagccggggc	agcgccgccc	gtaaacgccg	gtggcacgac	13680
aggcagcggg	gacagatgtg	ggacgatgag	gactccgccg	acgacagcag	cgtgttggac	13740

ttgggtggga	gtggtaaccc	gttcgctcac	ctgcgcccc	gtatcgggcg	catgatgtaa	13800
gagaaaccga	aaataaatga	tactcaccaa	ggccatggcg	accagcgtgc	gttcgtttct	13860
tctctgttgt	tgttgtatct	agtatgatga	ggcgtgcgta	cccggagggt	cctcctccct	13920
cgtacgagag	cgtgatgcag	caggcgatgg	cggcggcggc	gatgcagccc	ccgctggagg	13980
ctccttacgt	gccccgcgg	tacctggcgc	ctacggaggg	gcggaacagc	attcgttact	14040
cggagctggc	acccttgtac	gataccaccc	ggttgtacct	ggtggacaac	aagtcggcgg	14100
acatgcctc	gctgaactac	cagaacgacc	acagcaactt	cctgaccacc	gtggtgcaga	14160
acaatgactt	cacccccacg	gaggccagca	cccagaccat	caactttgac	gagcgtcgc	14220
ggtggggcgg	ccagctgaaa	accatcatgc	acaccaacat	gcccacgtg	aacgagttca	14280
tgtacagcaa	caagttcaag	gcgcgggtga	tggtctcccg	caagaccccc	aatgggggtga	14340
cagtacaga	ggattatgat	ggtagtcagg	atgagctgaa	gtatgaatgg	gtggaatttg	14400
agctgcccga	aggcaacttc	tcggtgacca	tgaccatcga	cctgatgaac	aacgccatca	14460
tcgacaatta	cttggcgggtg	gggcggcaga	acgggggtgct	ggagagcgac	atcggcgtga	14520
agttcgacac	taggaacttc	aggctgggct	gggaccccg	gaccgagctg	gtcatgccc	14580
gggtgtacac	caacgaggct	ttccatcccg	atattgtctt	gctgcccggc	tgcgggggtg	14640
acttcaccga	gagccgcctc	agcaacctgc	tgggcattcg	caagaggcag	cccttccagg	14700
aaggcttcca	gatcatgtac	gaggatctgg	aggggggcaa	catccccgcg	ctcctggatg	14760
tcgacgccta	tgagaaaagc	aaggaggatg	cagcagctga	agcaactgca	gccgtagcta	14820
ccgcctctac	cgaggtcagg	ggcgataatt	ttgcaagcgc	cgcagcagtg	gcagcggccg	14880
aggcggctga	aaccgaaagt	aagatagtca	ttcagccgg	ggagaaggat	agcaagaaca	14940
ggagctacaa	cgtactaccg	gacaagataa	acaccgccta	ccgcagctgg	tacctagcct	15000
acaactatgg	cgaccccag	aagggcgtgc	gctcctggac	gctgctcacc	acctcggacg	15060
tcacctgcgg	cgtggagcaa	gtctactgg	cgctgcccga	catgatgcaa	gacccggtca	15120
ccttccgctc	cacgcgtcaa	gttagcaact	acccgggtgg	gggcgccgag	ctcctgccc	15180
tctactccaa	gagcttcttc	aacgagcagg	ccgtctactc	gcagcagctg	cgcgcttca	15240
cctcgcttac	gcacgtcttc	aaccgcttcc	ccgagaacca	gatcctcgtc	cggccgccc	15300
cggccaccat	taccaccgtc	agtgaaaacg	ttcctgctct	cacagatcac	gggaccctgc	15360

cgctgcgag	cagtatccgg	ggagtccagc	gcgtgaccgt	tactgacgcc	agacgccgca	15420
cctgccccta	cgtctacaag	gccctgggca	tagtcgcgcc	gcgcgtcctc	tcgagccgca	15480
ccttctaaat	gtccattctc	atctcgccca	gtaataacac	cggttggggc	ctgcgcgcg	15540
ccagcaagat	gtacggaggc	gctcgccaac	gctccacgca	acaccccgtg	cgcgtgcgcg	15600
ggcacttccg	cgctccctgg	ggcgccctca	agggccgcgt	gcggtcgcg	accaccgtcg	15660
acgacgtgat	cgaccaggtg	gtggccgacg	cgcgcaacta	cacccccgcc	gccgcgcccc	15720
tctccaccgt	ggacgccgtc	atcgacagcg	tggtggccga	cgcgcgccgg	tacgccccgc	15780
ccaagagccg	gcggcggcg	atcgcccggc	ggcaccggag	cacccccgcc	atgcgcgcgg	15840
cgcgagcctt	gctgcgaggg	gccaggcgca	cgggacgcag	ggccatgctc	agggcggcca	15900
gacgcgcggc	ttcaggcgcc	agcgccggca	ggacccggag	acgcgcggcc	acggcggcgg	15960
cagcggccat	cgccagcatg	tcccgcccg	ggcgagggaa	cgtgtactgg	gtgcgcgacg	16020
ccgccaccgg	tgtgcgcggt	cccgtgcgca	ccgcccccc	tcgcacttga	agatgttcac	16080
ttcgcgatgt	tgatgtgtcc	cagcggcgag	gaggatgtcc	aagcgcaaat	tcaaggaaga	16140
gatgctccag	gtcatcgcg	ctgagatcta	cggccctgcg	gtggtgaagg	aggaaagaaa	16200
gccccgaaa	atcaagcggg	tcaaaaagga	caaaaaggaa	gaagaaagtg	atgtggacgg	16260
attggtggag	tttgtgcg	agttcgcccc	ccggcggcg	gtgcagtggc	gcgggcggaa	16320
ggtgcaaccg	gtgctgagac	ccggcaccac	cgtggtcttc	acgcccggcg	agcgctccgg	16380
caccgcttcc	aagcgctcct	acgacgaggt	gtacggggat	gatgatattc	tggagcaggc	16440
ggccgagcgc	ctgggcgagt	ttgcttacgg	caagcgcagc	cgttccgcac	cgaaggaaga	16500
ggcgggtgtcc	atcccgtgg	accacggcaa	ccccacgccg	agcctcaagc	ccgtgacctt	16560
gcagcaggtg	ctgccgaccg	cggcgccgcg	ccgggggttc	aagcgcgagg	gcgaggatct	16620
gtaccccacc	atgcagctga	tggtgccccaa	gcgccagaag	ctggaagacg	tgctggagac	16680
catgaaggtg	gacccggacg	tgcagcccga	ggtcaaggtg	cggcccatca	agcaggtggc	16740
cccgggcctg	ggcgtgcaga	ccgtggacat	caagattccc	acggagccca	tggaaacgca	16800
gaccgagccc	atgatcaagc	ccagcaccag	caccatggag	gtgcagacgg	atccctggat	16860
gccatcggct	cctagtcgaa	gaccccggcg	caagtacggc	gcggccagcc	tgctgatgcc	16920
caactacg	ctgcatacctt	ccatcatccc	cacgccgggc	taccgcggca	cgcgcttcta	16980

ccgcggtcat	accagcagcc	gccgccgcaa	gaccaccact	cgccgccgcc	gtcgccgcac	17040
cgccgctgca	accacccctg	ccgccctggg	gcggagagtg	taccgccgcg	gccgcgcacc	17100
tctgaccctg	ccgcgcgcgc	gctaccaccc	gagcatcgcc	atttaaactt	tcgcctgctt	17160
tgcatatcaa	tggccctcac	atgccgcctt	cgcggttccca	ttacgggcta	ccgaggaaga	17220
aaaccgcgcc	gtagaaggct	ggcgggggaa	gggatgcgtc	gccaccacca	ccggcggcgg	17280
cgcgccatca	gcaagcgggt	gggggggagg	ttcctgcccc	cgctgatccc	catcatcgcc	17340
gcggcgatcg	gggcgatccc	cggcattgct	tccgtggcgg	tgaggccctc	tcagcgccac	17400
tgagacacac	ttggaacat	cttgtaataa	accaatggac	tctgacgctc	ctggtcctgt	17460
gatgtgtttt	cgtagacaga	tggaagacat	caatttttcg	tccctggctc	cgcgacacgg	17520
cacgcggccg	ttcatgggca	cctggagcga	catcggcacc	agccaactga	acggggggcg	17580
cttcaattgg	agcagtctct	ggagcgggct	taagaatttc	gggtccacgc	ttaaaaccta	17640
tggcagcaag	gcgtggaaca	gcaccacagg	gcaggcgctg	agggataagc	tgaaagagca	17700
gaacttccag	cagaaggtgg	tcgatgggct	cgccctgggc	atcaacgggg	tggtggacct	17760
ggccaaccag	gccgtgcagc	ggcagatcaa	cagccgcctg	gaccgggtgc	cgcccgccgg	17820
ctccgtggag	atgccgcagg	tggaggagga	gctgcctccc	ctggacaagc	ggggcgagaa	17880
gcgaccccg	cccgatgcgg	aggagacgct	gctgacgcac	acggacgagc	cgccccgta	17940
cgaggaggcg	gtgaaactgg	gtctgcccac	cacgcggccc	atcgcgcccc	tggccaccgg	18000
ggtgctgaaa	cccgaagc	ccgcgaccct	ggacttgctt	cctccccagc	cttcccggcc	18060
ctctacagt	gctaagcccc	tgccgccggg	ggccgtggcc	cgcgcgcgac	ccggggggcac	18120
cgcccgccct	catgcgaact	ggcagagcac	tctgaacagc	atcgtagggtc	tgggagtgc	18180
gagtgtgaag	cgccgccgct	gctattaaac	ctaccgtagc	gcttaacttg	cttgtctgtg	18240
tgtgtatgta	ttatgtcgcc	gccgccgctg	tccaccagaa	ggaggagtga	agaggcgcgt	18300
cgccgagttg	caagatggcc	accccatcga	tgctgcccc	gtgggcgtac	atgcacatcg	18360
ccggacagga	cgcttcggag	tacctgagtc	cgggtctggg	gcagtttgcc	cgcgccacag	18420
acacctactt	cagtctgggg	aacaagttaa	ggaacccac	ggtggcgccc	acgcacgatg	18480
tgaccaccga	ccgcagccag	cggctgacgc	tgcgcttcgt	gcccgtggac	cgcgaggaca	18540
acacctactc	gtacaaagt	cgctacacgc	tggccgtggg	cgacaaccgc	gtgctggaca	18600

tggccagcac	ctactttgac	atccgcggcg	tgctggatcg	gggccctagc	ttcaaaccct	18660
actccggcac	cgcttacaac	agtctggccc	ccaagggagc	acccaacact	tgtcagtgga	18720
catataaagc	cgatggtgaa	actgccacag	aaaaaaccta	tacatatgga	aatgcacccg	18780
tgcagggcat	taacatcaca	aaagatggta	ttcaacttgg	aactgacacc	gatgatcagc	18840
caatctacgc	agataaaacc	tatcagcctg	aacctcaagt	gggtgatgct	gaatggcatg	18900
acatcactgg	tactgatgaa	aagtatggag	gcagagctct	taagcctgat	acaaaaatga	18960
agccttgтта	tggttctttt	gccaagccta	ctaataaaga	aggaggctcag	gcaaattgta	19020
aaacaggaac	aggcactact	aaagaatatg	acatagacat	ggcttttcttt	gacaacagaa	19080
gtgcggctgc	tgctggccta	gctccagaaa	ttgttttgta	tactgaaaat	gtggatttgg	19140
aaactccaga	tacccatatt	gtatacaaag	caggcacaga	tgacagcagc	tcttctatta	19200
atttgggtca	gcaagccatg	cccaacagac	ctaactacat	tggttttcaga	gacaacttta	19260
tcgggctcat	gtactacaac	agcactggca	atatgggggt	gctggccggt	caggcttctc	19320
agctgaatgc	tgtggttgac	ttgcaagaca	gaaacaccga	gctgtcctac	cagctcttgc	19380
ttgactctct	gggtgacaga	acccggtatt	tcagtatgtg	gaatcaggcg	gtggacagct	19440
atgatcctga	tgtgcgcat	attgaaaatc	atgggtgtgga	ggatgaactt	cccaactatt	19500
gtttccctct	ggatgctggt	ggcagaacag	atacttatca	gggaattaag	gctaattgaa	19560
ctgatcaaac	cacatggacc	aaagatgaca	gtgtcaatga	tgctaattgag	ataggcaagg	19620
gtaatccatt	cgccatggaa	atcaacatcc	aagccaacct	gtggagggaac	ttcctctacg	19680
ccaacgtggc	cctgtacctg	cccgactctt	acaagtacac	gccggccaat	gttaccctgc	19740
ccaccaacac	caacacctac	gattacatga	acggccgggt	ggtaggcgcc	tcgctggtgg	19800
actcctacat	caacatcggt	gcgcgctggt	cgctggatcc	catggacaac	gtgaaccctt	19860
tcaaccacca	ccgcaatgcg	gggctgcgct	accgctccat	gctcctgggc	aacgggcgct	19920
acgtgccctt	ccacatccag	gtgccccaga	aatttttcgc	catcaagagc	ctcctgctcc	19980
tgcccgggtc	ctacacctac	gagtgggaact	tccgcaagga	cgtcaacatg	atcctgcaga	20040
gctccctcgg	caacgacctg	cgcacggacg	gggcctccat	ctccttcacc	agcatcaacc	20100
tctacgccac	cttcttcccc	atggcgcaca	acacggcctc	cacgctcgag	gcatgctgc	20160
gcaacgacac	caacgaccag	tccttcaacg	actacctctc	ggcggccaac	atgctctacc	20220

ccatccccggc caacgccacc aacgtgcccc tctccatccc ctcgcgcaac tgggccgcct	20280
tccgcgggctg gtccttcacg cgtctcaaga ccaaggagac gccctcgctg ggctccgggt	20340
tcgacccta cttcgtctac tcgggctcca tcccctacct cgacggcacc ttctacctca	20400
accacacctt caagaaggct tccatcacct tcgactcctc cgtcagctgg cccggcaacg	20460
accggctcct gacgcccac gagttcgaaa tcaagcgac cgtcgacggc gagggctaca	20520
acgtggccca gtgcaacatg accaaggact ggttcctggg ccagatgctg gccactaca	20580
acatcggcta ccagggttc tacgtgccc agggctacaa ggaccgcatg tactccttct	20640
tccgcaactt ccagcccatg agccgccagg tggtagcga ggtcaactac aaggactacc	20700
aggccgtcac cctggcctac cagcacaaca actcgggctt cgtaggctac ctcgccccca	20760
ccatgcgcca gggccagccc taccgccca actacccta cccgctcatc ggcaagagcg	20820
ccgtcaccag cgtcaccag aaaaagttcc tctgcgacag ggcatgtgg cgcatcccct	20880
tctccagcaa cttcatgtcc atgggcgcgc tcaccgacct cggccagaac atgctctatg	20940
ccaactccgc ccacgcgcta gacatgaatt tcgaagtcga ccccatggat gagtccaccc	21000
ttctctatgt tgtcttcgaa gtcttcgacg tcgtccgagt gcaccagccc caccgcggcg	21060
tcatcgaggc cgtctacctg cgcaccccct tctcggccgg taacgccacc acctaagctc	21120
ttgcttcttg caagccatgg ccgcgggctc cggcgagcag gagctcaggg ccatcatccg	21180
cgacctgggc tgcgggcctt acttctctggg caccttcgat aagcgcttcc cgggattcat	21240
ggccccgcac aagctggcct gcgccatcgt caacacggcc ggccgcgaga ccggggggcga	21300
gcactggctg gccttcgcct ggaacccgcg ctcgaacacc tgctacctct tcgaccctt	21360
cgggttctcg gacgagcgcc tcaagcagat ctaccagttc gagtacgagg gcctgctgcg	21420
ccgcagcgcc ctggccaccg aggaccgctg cgtcaccctg gaaaagtcca cccagaccgt	21480
gcagggtccg cgctcggccg cctgcgggct cttctgctgc atgttcctgc acgccttcgt	21540
gcactggccc gaccgcccc tggacaagaa cccaccatg aacttgctga cgggggtgcc	21600
caacggcatg ctccagtcgc cccaggtgga acccaccctg cgccgcaacc aggaggcgct	21660
ctaccgcttc ctcaactccc actccgccta ctttcgctcc caccgcgcgc gcatcgagaa	21720
ggccaccgcc ttcgaccgca tgaatcaaga catgtaaacc gtgtgtgtat gttaaattgc	21780
tttaataaac agcactttca tgttacacat gcatctgaga tgatttatatt agaaatcgaa	21840

agggttctgc cgggtctcgg catggcccgc gggcagggac acgttgcgga actggtactt	21900
ggccagccac ttgaactcgg ggatcagcag tttgggcagc ggggtgtcgg ggaaggagtc	21960
ggtccacagc ttccgcgtca gttgcagggc gccagcagg tcgggcgcgg agatcttgaa	22020
atcgcatgtt ggacccgcgt tctgcgcgcg ggagttgcgg tacacggggg tgcagcactg	22080
gaacaccatc agggccgggt gcttcacgct cgccagcacc gtcgcgtcgg tgatgctctc	22140
cacgtcgagg tcctcggcgt tggccatccc gaaggggggtc atcttgtagg tctgccttcc	22200
catggtgggc acgcacccgg gcttgtgggt gcaatcgtag tgcaggggga tcagcatcat	22260
ctgggcctgg tcggcggtta tccccgggta catggccttc atgaaagcct ccaattgcct	22320
gaacgcctgc tgggccttgg ctccctcggg gaagaagacc ccgcaggact tgctagagaa	22380
ctggttggtg gcgcacccgg cgtcgtgcac gcagcagcgc gcgtcgttgt tggccagctg	22440
caccacgctg cgcggccagc ggttctgggt gatcttggcc cggtcggggg tctccttcag	22500
cgcgcgctgc ccgttctcgc tcgccacatc catctcgatc atgtgctcct tctggatcat	22560
ggtggtcccg tgcaggcacc gcagcttgcc ctccgacctg gtgcacccgt gcagccacag	22620
cgcgcacccg gtgcactccc agttcttgtg ggcgatctgg gaatgcgcgt gcacgaagcc	22680
ctgcaggaag cggcccatca tgggtggtcag ggtcttggtt ctagtgaagg tcagcggaat	22740
gccgcggtgc tcctcgttga tgtacaggtg gcagatgcgg cggtagacac ctgcctgctc	22800
gggcatcagc tggaagttgg ctttcaggtc ggtctccacg cggtagcggg ccatcagcat	22860
agtcatgatt tccataccct tctcccaggc cgagacgatg ggcaggctca tagggttctt	22920
caccatcatc ttagcgctag cagccgcggc caggggggtc ctctcgtcca ggtctcaaaa	22980
gctccgcttg ccgtccttct cgggtgatccg caccgggggg tagctgaagc ccacggccgc	23040
cagctcctcc tcggcctgtc tttcgtcctc gctgtcctgg ctgacgtcct gcaggaccac	23100
atgcttggtc ttgcgggggt tcttcttggg cggcagcggc ggcggagatg ttggagatgg	23160
cgagggggag cgcgagttct cgctcaccac tactatctct tcctcttctt ggtccgaggc	23220
cacgcggcgg taggtatgtc tcttcggggg cagaggcgga ggcgacgggc tctcgccgcc	23280
gcgacttggc ggatggctgg cagagcccct tccgcgttcg ggggtgcgct cccggcggcg	23340
ctctgactga cttcctccgc ggccggccat tgtgttctcc tagggaggaa caacaagcat	23400
ggagactcag ccatcgccaa cctcgccatc tgccccacc gccgacgaga agcagcagca	23460

gcagaatgaa agcttaaccg ccccgccgcc cagccccgcc acctccgacg cggccgtccc	23520
agacatgcaa gagatggagg aatccatcga gattgacctg ggctatgtga cggccgcgga	23580
gcacgaggag gagctggcag tgcgcttttc acaagaagag atacaccaag aacagccaga	23640
gcaggaagca gagaatgagc agagtcaggc tgggctcgag catgacggcg actacctcca	23700
cctgagcggg ggggaggacg cgctcatcaa gcatctggcc cggcaggcca ccatcgtcaa	23760
ggatgcgctg ctcgaccgca ccgaggtgcc cctcagcgtg gaggagctca gccgcgccta	23820
cgagttgaac ctcttctcgc cgcgctgcc cccaagcgc cagcccaatg gcacctgcga	23880
gccaacccg cgctcaact tctaccggt cttcgcggtg cccgaggccc tggccaccta	23940
ccacatcttt ttcaagaacc aaaagatccc cgtctcctgc cgcgccaacc gcacccgcgc	24000
cgacgccctt ttcaacctgg gtcccggcgc ccgcctacct gatatcgctt ccttggaaga	24060
ggttccaag atcttcgagg gtctgggcag cgacgagact cgggccgcga acgctctgca	24120
aggagaagga ggagagcatg agcaccacag cgccctggtc gagttggaag gcgacaacgc	24180
gcggctggcg gtgctcaaac gcacggtcga gctgacctat ttcgcctacc cggctctgaa	24240
cctgcccccc aaagtcatga gcgcggtcat ggaccaggtg ctcatcaagc gcgcgtcgcc	24300
catctccgag gacgagggca tgcaagactc cgaggagggc aagcccgtgg tcagcgacga	24360
gcagctggcc cgggtggctgg gtcctaattgc tagtccccag agtttggaag agcggcgcaa	24420
actcatgatg gccgtggtcc tgggtgacct ggagctggag tgcctgcgcc gcttcttcgc	24480
cgacgaggag accctgcga aggtcgagga gaacctgcac tacctcttca ggcacggggt	24540
cgtgcgccag gcctgcaaga tctccaacgt ggagctgacc aacctgggtt cctacatggg	24600
catcttgca gagaaccgcc tggggcagaa cgtgctgcac accaccctgc gcggggaggc	24660
ccggcgcgac tacatccgcg actgcgtcta cctctacctc tgccacacct ggcagacggg	24720
catgggcgtg tggcagcagt gtctggagga gcagaacctg aaagagctct gcaagctcct	24780
gcagaagaac ctcaagggtc tgtggaccgg gttcgacgag cgcaccaccg cctcggacct	24840
ggccgacctc attttccccg agcgcctcag gctgacgctg cgcaacggcc tgcccgactt	24900
tatgagccaa agcatgttgc aaaactttcg ctctttcatc ctggaacgct ccggaatcct	24960
gcccgccacc tgctccgcgc tgccctcgga cttcgtgccg ctgaccttcc gcgagtgcc	25020
cccgccgctg tggagccact gctacctgct gcgcctggcc aactacctgg cctaccactc	25080

ggacgtgatac	gaggacgtca	gcggcgaggg	cctgctcgag	tgccactgcc	gctgcaacct	25140
ctgcacgccg	caccgctccc	tggcctgcaa	ccccagctg	ctgagcgaga	cccagatcat	25200
cggcaccttc	gagttgcaag	ggcccagcga	aggcgagggt	tcagccgcca	aggggggtct	25260
gaaactcacc	ccggggctgt	ggacctcggc	ctacttgctc	aagttcgtgc	ccgaggacta	25320
ccatcccctt	gagatcaggt	tctacgagga	ccaatcccat	ccgcccagg	ccgagctgtc	25380
ggcctgcgtc	atcaccagg	gggcgatcct	ggcccaattg	caagccatcc	agaaatccc	25440
ccaagaattc	ttgctgaaaa	agggccgcgg	ggtctacctc	gacccccaga	ccggtgagga	25500
gctcaacccc	ggcttcccc	aggatgcccc	gaggaaacaa	gaagctgaaa	gtggagctgc	25560
cggccgtgga	ggatttggag	gaagactggg	agaacagcag	tcaggcagag	gaggaggaga	25620
tggaggaaga	ctgggacagc	actcaggcag	aggaggacag	cctgcaagac	agtctggagg	25680
aagacgagga	ggaggcagag	gaggaggtgg	aagaagcagc	cggccgcca	ccgtcgtcct	25740
cggcggggga	gaaagcaagc	agcacggata	ccatctccgc	tccgggtcgg	ggtcccgtc	25800
gaccacacag	tagatgggac	gagaccggac	gattccccga	ccccaccacc	cagaccggt	25860
agaaggagcg	gcagggatac	aagtcctggc	gggggcacaa	aaacgccatc	gtctcctgct	25920
tgcaggcctg	cgggggcaac	atctccttca	cccggcgcta	cctgctcttc	caccgcgggg	25980
tgaactttcc	ccgcaacatc	ttgcattact	accgtcacct	ccacagcccc	tactacttcc	26040
aagaagaggc	agcagcagca	gaaaaagacc	agcagaaaac	cagcagctag	aaaatccaca	26100
gcggcggcag	caggtggact	gaggatcgcg	gcgaacgagc	cggcgcaaac	ccgggagctg	26160
aggaaccgga	tctttccac	cctctatgcc	atcttccagc	agagtcgggg	gcaggagcag	26220
gaactgaaag	tcaagaaccg	ttctctgcgc	tcgctcacc	gcagttgtct	gtatcacaag	26280
agcgaagacc	aacttcagcg	cactctcgag	gacgccgagg	ctctcttcaa	caagtactgc	26340
gcgctcactc	ttaaagagta	gcccgcgccc	gccagtcgc	agaaaaaggc	gggaattacg	26400
tcacctgtgc	ccttcgccct	agccgcctcc	acccatcatc	atgagcaaag	agattccac	26460
gccttacatg	tggagctacc	agccccagat	gggcctggcc	gccggtgccg	cccaggacta	26520
ctccaccgcg	atgaattggc	tcagcgccgg	gcccgcgatg	atctcacggg	tgaatgacat	26580
ccgcgcccac	cgaaccaga	tactcctaga	acagtcagcg	ctcaccgcca	cgcggcgcaa	26640
tcacctcaat	ccgcgtaatt	ggcccgcgcg	cctggtgtac	caggaaattc	cccagcccac	26700

gaccgtacta	cttccgcgag	acgcccaggc	cgaagtccag	ctgactaact	caggtgtcca	26760
gctggcgggc	ggcgccaccc	tgtgtcgtca	ccgccccgct	cagggtataa	agcggctggg	26820
gatccggggc	agaggcacac	agctcaacga	cgaggtgggtg	agctcttcgc	tgggtctgcg	26880
acctgacgga	gtcttccaac	tcgccggatc	ggggagatct	tccttcacgc	ctcgtcaggc	26940
cgtcctgact	ttggagagtt	cgtcctcgca	gccccgctcg	ggtagcatcg	gcactctcca	27000
gttcgtggag	gagttcactc	cctcgggtcta	cttcaacccc	ttctccggct	ccccggcca	27060
ctacccggac	gagttcatcc	cgaacttcga	cgccatcagc	gagtcgggtgg	acggctacga	27120
ttgaatgtcc	catggtggcg	cagctgacct	agctcggctt	cgacacctgg	accactgccg	27180
ccgcttccgc	tgcttcgctc	gggatctcgc	cgagtttgcc	tactttgagc	tgcccgagga	27240
gcaccctcag	ggccccggcc	acggagtgcg	gatcgtcgtc	gaagggggcc	tcgactccca	27300
cctgcttcgg	atcttcagcc	agcgtccgat	cctggtcgag	cgcgagcaag	gacagaccct	27360
tctgactctg	tactgcatct	gcaaccaccc	cggcctgcat	gaaagtcttt	gttgtctgct	27420
gtgtactgag	tataataaaa	gctgagatca	gcgactactc	cggacttccg	tgtgttcctg	27480
aatccatcaa	ccagtctttg	ttcttcaccg	ggaacgagac	cgagctccag	ctccagtgtg	27540
agccccacaa	gaagtacctc	acctggctgt	tccagggctc	cccgatcgcc	gttgtcaacc	27600
actgcgacaa	cgacggagtc	ctgctgagcg	gccctgccaa	ccttactttt	tccacccgca	27660
gaagcaagct	ccagctcttc	caacccttcc	tccccgggac	ctatcagtgc	gtctcgggac	27720
cctgccatca	caccttcac	ctgatcccga	ataccacagc	gtcgctcccc	gtactaaca	27780
acaaaactaa	cctccaccaa	cgccaccgtc	gcgacctttc	tgaatcta	actaccaccc	27840
acaccggagg	tgagctccga	ggtcaaccaa	cctctgggat	ttactacggc	ccctgggagg	27900
tggttgggtt	aatagcgcta	ggcctagttg	cgggtgggct	tttggttctc	tgctacctat	27960
acctcccttg	ctgttcgtac	ttagtggtgc	tgtgttgctg	gtttaagaaa	tggggaagat	28020
caccctagtg	agctgcggtg	cgctgggtggc	ggtgttgctt	tcgattgtgg	gactgggcgg	28080
tgcggtgta	gtgaaggaga	aggccgatcc	ctgcttgcat	ttcaatccca	acaaatgcca	28140
gctgagtttt	cagcccgatg	gcaatcgggtg	cgcggtactg	atcaagtgcg	gatgggaatg	28200
cgagaacgtg	agaatcgagt	acaataacaa	gactcggaac	aatactctcg	cgtccgtgtg	28260
gcagcccggg	gaccccgagt	ggtacaccgt	ctctgtcccc	ggtgctgacg	gctccccgcg	28320

caccgtgaat aatacttttca tttttgcgca catgtgcgac acggtcatgt ggatgagcaa	28380
gcagtacgat atgtggcccc ccacgaagga gaacatcgtg gtcttctcca tcgcttacag	28440
cctgtgcacg gcgctaataca ccgctatcgt gtgcctgagc attcacatgc tcatcgctat	28500
tcgccccaga aataatgccg aaaaagaaaa acagccataa cgttttttttt cacacctttt	28560
tcagaccatg gcctctgtta aatttttgct tttatttgcc agtctcattg ccgtcattca	28620
tggaatgagt aatgagaaaa ttactattta cactggcact aatcacacat tgaaagggtcc	28680
agaaaaagcc acagaagttt catgggtattg ttattttaat gaatcagatg tatctactga	28740
actctgtgga aacaataaca aaaaaaatga gagcattact ctcatcaagt ttcaatgtgg	28800
atctgactta accctaatta acatcactag agactatgta ggtatgtatt atggaactac	28860
agcaggcatt tcggacatgg aattttatca agtttctgtg tctgaaccca ccacgcctag	28920
aatgaccaca accacaaaaa ctacacctgt taccactatg cagctcacta ccaataacat	28980
ttttgccatg cgtcaaatgg tcaacaatag cactcaaccc accccaccca gtgaggaaat	29040
tcccaaattc atgattggca ttattgttgc tgtagtggtg tgcatgttga tcatcgctt	29100
gtgcatgggtg tactatgcct tctgctacag aaagcacaga ctgaacgaca agctggaaca	29160
cttactaagt gttgaatttt aattttttag aaccatgaag atcctaggcc ttttaatttt	29220
ttctatcatt acctctgctc tatgcaattc tgacaatgag gacgttactg tcgttgtcgg	29280
atcaaattat acactgaaag gtccagcgaa gggtatgctt tcgtgggtatt gctattttgg	29340
atctgacact acagaaactg aattatgcaa tcttaagaat ggcaaaattc aaaattctaa	29400
aattaacaat tatatatgca atggtactga tctgatactc ctcaatatca cgaaatcata	29460
tgctggcagt tacacctgcc ctggagatga tgctgacagt atgatTTTTT acaaagtaac	29520
tgttgttgat cccactactc cacctccacc caccacaact actcacacca cacacacaga	29580
tcaaaccgca gcagaggagg cagcaaagtt agccttgtag gtccaagaca gttcatttgt	29640
tggcattacc cctacacctg atcagcgggtg tccggggctg ctagtcagcg gcattgtcgg	29700
tgtgctttcg ggattagcag tcataatcat ctgcatgttc atttttgctt gctgctatag	29760
aaggctttac cgacaaaaat cagaccact gctgaacctc tatgtttaat tttttccaga	29820
gtcatgaagg cagttagcgc tctagttttt tgttctttga ttggcattgt tttttgcaat	29880
cctattccta aagttagctt tattaaagat gtgaatgtta ctgagggggg caatgtgaca	29940

ctggtaggtg tagagggtgc tgaaaacacc acctggacaa aataccacct caatgggtgg	30000
aaagatattt gcaattggag tgtattagtt tatacatgtg agggagttaa tcttaccatt	30060
gtcaatgcca cctcagctca aaatggtaga attcaaggac aaagtgtcag tgtatctaata	30120
gggtatttta cccaacatac ttttatctat gacgttaaag tcataccact gcctacgcct	30180
agcccaccta gcactaccac acagacaacc cacactacac agacaaccac atacagtaca	30240
ttaaatacagc ctaccaccac tacagcagca gaggttgcca gctcgtctgg ggtccgagtg	30300
gcatttttga tgtgggcccc atctagcagt cccactgcta gtaccaatga gcagactact	30360
gaatttttgt ccactgtcga gagccacacc acagctacct ccagtgcctt ctctagcacc	30420
gccaatctct cctcgctttc ctctacacca atcagtcccg ctactactcc tagccccgct	30480
cctcttccca ctcccctgaa gcaaacagac ggcggcatgc aatggcagat caccctgctc	30540
attgtgatcg ggttgggtcat cctggccgtg ttgctctact acatcttctg ccgccgcatt	30600
cccaacgcgc accgcaagcc ggtctacaag cccatcattg tcgggcagcc ggagccgctt	30660
caggtggaag ggggtctaag gaatcttctc ttctctttta cagtatggtg attgaactat	30720
gattcctaga caattcttga tcactattct tatctgcctc ctccaagtct gtgccaccct	30780
cgctctggtg gccaacgcca gtccagactg tattgggccc ttgcctcct acgtgctctt	30840
tgcttccacc acctgcatct gctgctgtag catagtctgc ctgcttatca ccttcttcca	30900
gttcattgac tggatctttg tgcgcatcgc ctacctgcgc caccaccccc agtaccgcga	30960
ccagcgagtg gcgcggctgc tcaggctcct ctgataagca tgcgggctct gctacttctc	31020
gcgcttctgc tgttagtgtc cccccgtccc gtcgaccccc ggtccccac ccagtcccc	31080
gaggaggtcc gcaaatgcaa attccaagaa ccctggaaat tcctcaaata ctaccgcaa	31140
aaatcagaca tgcattcccag ctggatcatg atcattggga tcgtgaacat tctggcctgc	31200
accctcatct cttttgtgat ttaccctgc tttgactttg gttggaactc gccagaggcg	31260
ctctatctcc cgcctgaacc tgacacacca ccacagcaac ctcaaggaca cgcactacca	31320
ccactacagc ctaggccaca atacatgccc atattagact atgaggccga gccacagcga	31380
cccatgctcc ccgctattag ttacttcaat ctaaccggcg gagatgactg acccactggc	31440
caacaacaac gtcaacgacc ttctcctgga catggacggc cgcgcctcgg agcagcgact	31500
cgcccaactt cgcattcgcc agcagcagga gagagccgtc aaggagctgc aggatgcggt	31560

ggccatccac cagtgaaga gaggcattctt ctgcctgggtg aaacaggcca agatctccta	31620
cgaggtcact ccaaacgacc atcgctcttc ctacgagctc ctgcagcagc gccagaagtt	31680
cacctgcctg gtcggagtca accccatcgt catcaccag cagtctggcg ataccaaggg	31740
gtgcatccac tgctcctgcg actccccga ctgcgtccac actctgatca agaccctctg	31800
cggcctccgc gacctcctcc ccatgaacta atcacccct tatccagtga aataaagatc	31860
atattgatga tgattttaca gaaataaaaa ataatcattt gatttgaaat aaagatacaa	31920
tcatattgat gatttgagtt taacaaaaaa ataaagaatc acttacttga aatctgatac	31980
caggctcttg tccatgtttt ctgccaacac cacttcactc ccctcttccc agctctggta	32040
ctgcaggccc cggcgggctg caaacttcct ccacacgctg aaggggatgt caaattcctc	32100
ctgtccctca atcttcattt tatcttctat cagatgtcca aaaagcgcgt ccgggtggat	32160
gatgacttcg acccgtcta cccctacgat gcagacaacg caccgaccgt gcccttcac	32220
aacccccct tcgtctcttc agatggattc caagagaagc ccctgggggt gttgtccctg	32280
cgactggccg acccgtcac caccaagaac ggggaaatca ccctcaagct gggagagggg	32340
gtggacctcg attcctcggg aaaactcatc tccaacacgg ccaccaaggc cgccgccct	32400
ctcagttttt ccaacaacac catttcctt aacatggatc acccctttta cactaaagat	32460
ggaaaattat cttacaagt ttctccacca ttaaataac tgagaacaag cattctaaac	32520
acactagctt taggttttgg atcaggttta ggactccgtg gctctgcctt ggcagtacag	32580
ttagtctctc cacttacatt tgatactgat ggaaacataa agcttacctt agacagaggt	32640
ttgcatgtta caacaggaga tgcaattgaa agcaacataa gctgggctaa aggtttaaaa	32700
tttgaagatg gagccatagc aaccaacatt ggaaatgggt tagagtttgg aagcagtagt	32760
acagaaacag gtgttgatga tgcttaccca atccaagtta aacttggatc tggccttagc	32820
tttgacagta caggagccat aatggctggt aacaaagaag acgataaact cactttgtgg	32880
acaacacctg atccatcacc aaactgtcaa atactcgcag aaaatgatgc aaaactaaca	32940
ctttgcttga ctaaattgtg tagtcaaata ctggccactg tgtcagtctt agttgtagga	33000
agtggaaacc taaacccat tactggcacc gtaagcagtg ctcaggtgtt tctacgtttt	33060
gatgcaaacg gtgttctttt aacagaacat tctacactaa aaaaatactg ggggtatagg	33120
caggagata gcatagatgg cactccatat accaatgctg taggattcat gcccaattta	33180

aaagcttatc	caaagtcaca	aagttctact	actaaaaata	atatagtagg	gcaagtatac	33240
atgaatggag	atgtttcaaa	acctatgctt	ctcactataa	ccctcaatgg	tactgatgac	33300
agcaacagta	catattcaat	gtcattttca	tacacctgga	ctaatggaag	ctatgttgga	33360
gcaacatttg	gggctaactc	ttataccttc	tcatacatcg	cccaagaatg	aacactgtat	33420
cccaccctgc	atgccaaccc	ttccccaccc	actctgtgga	acaaactctg	aaacacaaaa	33480
taaaataaag	ttcaagtgtt	ttattgattc	aacagtttta	caggattcga	gcagttatth	33540
ttcctccacc	ctcccaggac	atggaataca	ccaccctctc	cccccgca	gccttgaaca	33600
tctgaatgcc	attggtgatg	gacatgcttt	tggtctccac	gttccacaca	gtttcagagc	33660
gagccagtct	cgggtcggtc	agggagatga	aaccctccgg	gcactcccgc	atctgcacct	33720
cacagctcaa	cagctgagga	ttgtcctcgg	tggtcgggat	cacggttatc	tggaagaagc	33780
agaagagcgg	cgggtgggaat	catagtccgc	gaacgggatc	ggccgggtgt	gtcgcatcag	33840
gccccgcagc	agtcgctgcc	gccgccgctc	cgtcaagctg	ctgctcaggg	ggtccgggtc	33900
cagggactcc	ctcagcatga	tgcccacggc	cctcagcatc	agtcgtctgg	tgcggcgggc	33960
gcagcagcgc	atgcggatct	cgctcaggtc	gctgcagtac	gtgcaacaca	gaaccaccag	34020
gttgttcaac	agtccatagt	tcaacacgct	ccagccgaaa	ctcatcgagg	gaaggatgct	34080
accacagtgg	ccgtcgtacc	agatcctcag	gtaaatcaag	tggtgcccc	tccagaacac	34140
gctgcccacg	tacatgatct	ccttgggcat	gtggcggttc	accacctccc	ggtaccacat	34200
cacctcttgg	ttgaacatgc	agccccggat	gatcctgcgg	aaccacaggg	ccagcaccgc	34260
ccgccccgcc	atgcagcgaa	gagaccccgg	gtcccggcaa	tggcaatgga	ggaccaccg	34320
ctcgtacccg	tggatcatct	gggagctgaa	caagtctatg	ttggcacagc	acaggcatat	34380
gctcatgcat	ctcttcagca	ctctcaactc	ctcgggggtc	aaaaccatat	cccagggcac	34440
ggggaactct	tgcaggacag	cgaaccccgc	agaacagggc	aatcctcgca	cagaacttac	34500
attgtgcatg	gacagggtat	cgcaatcagg	cagcaccggg	tgatcctcca	ccagagaagc	34560
gcgggtctcg	gtctcctcac	agcgtggtaa	gggggcccgg	cgatacgggt	gatggcggga	34620
cgcggtgat	cgtgttcgcg	accgtgtcat	gatgcagttg	ctttcggaca	ttttcgtact	34680
tgctgtagca	gaacctggtc	cgggcgctgc	acaccgatcg	ccggcggcgg	tctcggcgct	34740
tggaacgctc	ggtgttgaaa	ttgtaaaaca	gccactctct	cagaccgtgc	agcagatcta	34800

gggcctcagg	agtgatgaag	atcccatcat	gcctgatggc	tctgatcaca	tcgaccaccg	34860
tggaatgggc	cagaccagc	cagatgatgc	aattttgttg	ggtttcggtg	acggcggggg	34920
aggaagaac	aggaagaacc	atgattaact	tttaatccaa	acggtctcgg	agtacttcaa	34980
aatgaagatc	gcggagatgg	cacctctcgc	ccccgctgtg	ttggtggaaa	ataacagcca	35040
ggtcaaaggt	gatacggttc	tcgagatggt	ccacgggtggc	ttccagcaaa	gcctccacgc	35100
gcacatccag	aaacaagaca	atagcgaaag	cgggagggtt	ctctaattcc	tcaatcatca	35160
tgttacactc	ctgcaccatc	cccagataat	tttcattttt	ccagccttga	atgattcgaa	35220
ctagtctctg	aggtaaatcc	aagccagcca	tgataaagag	ctcgcgcaga	gcgccctcca	35280
ccggcattct	taagcacacc	ctcataattc	caagatatct	tgctcctggt	tcacctgcag	35340
cagattgaca	agcggaatat	caaaatctct	gccgcgatcc	ctgagctcct	ccctcagcaa	35400
taactgtaag	tactctttca	tatcctctcc	gaaattttta	gccataggac	caccaggaat	35460
aagattaggg	caagccacag	tacagataaa	ccgaagtcct	ccccagttag	cattgcaaaa	35520
tgcaagactg	ctataagcat	gctggctaga	cccggtgata	tcttccagat	aactggacag	35580
aaaatcgccc	aggcaatttt	taagaaaatc	aacaaaagaa	aaatcctcca	ggtggacgtt	35640
tagagcctcg	ggaacaacga	tgaagtaa	gcaagcggtg	cggtccagca	tggttagtta	35700
gctgatctgt	agaaaaaaca	aaaatgaaca	ttaaaccatg	ctagcctggc	gaacagggtg	35760
gtaaatcggt	ctctccagca	ccaggcaggc	cacgggggtct	ccggcgcgac	cctcgtaaaa	35820
attgtcgcta	tgattgaaaa	ccatcacaga	gagacgttcc	cggtggccgg	cgtgaatgat	35880
tcgacaagat	gaatacacc	ccggaacatt	ggcgtccgcg	agtgaaaaaa	agcgcccgag	35940
gaagcaataa	ggcactacaa	tgctcagtct	caagtccagc	aaagcgatgc	catgcggatg	36000
aagcacaaaa	ttctcagggt	cgtacaaaat	gtaattactc	ccctcctgca	caggcagcaa	36060
agccccgat	ccctccaggt	acacatacaa	agcctcagcg	tccatagctt	accgagcagc	36120
agcacacaac	aggcgcaaga	gtcagagaaa	ggctgagctc	taacctgtcc	acccgctctc	36180
tgctcaatat	atagcccaga	tctacactga	cgtaaaggcc	aaagtctaaa	aatacccgcc	36240
aaataatcac	acacgcccag	cacacgccc	gaaaccggtg	acacactcaa	aaaaatacgc	36300
gcatttctc	aaacgccc	aactgccgtc	atttccgggt	tcccacgcta	cgtcatcaaa	36360
acacgacttt	caaattccgt	cgaccgttaa	aaacgtcacc	cgccccgccc	ctaacggctg	36420

cccgctctctc agccaatcag cgccccgcat ccccaaattc aaacacctca tttgcatatt	36480
aacgcgcaca aaaagtttga ggtatattat tgatgatgg	36519

<210> 2
 <211> 31588
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence: Synthetic
 polynucleotide

<400> 2	
ccatcttcaa taatatacct caaacttttt gtgcgcgtta atatgcaaat gaggcgtttg	60
aatttgggga ggaagggcgg tgattggtcg agggatgagc gaccgttagg ggcggggcga	120
gtgacgtttt gatgacgtgg ttgcgaggag gagccagttt gcaagttctc gtgggaaaag	180
tgacgtcaaa cgaggtgtgg tttgaacacg gaaataactca attttcccgc gctctctgac	240
aggaaatgag gtgtttctgg gcggatgcaa gtgaaaacgg gccattttctg cgcgaaaact	300
gaatgaggaa gtgaaaatct gagtaatttc gcgtttatgg caggaggagg tatttgccga	360
gggccgagta gactttgacc gattacgtgg gggtttcgat taccgtgttt ttcacctaaa	420
tttccgcgta cgggtgtcaaa gtccggtgtt tttacgtagg tgtcagctga tcgccagggt	480
atttaaactt gcgctctcca gtcaagaggc cactcttgag tgccagcgag aagagttttc	540
tcctccgcgc cgcgagtcag atctacactt tgaaagtagg gataacaggg taatgacatt	600
gattattgac tagttgttaa tagtaatcaa ttacggggtc attagttcat agcccatata	660
tggagttccg cgttacataa cttacggtaa atggcccgc tggtctgaccg cccaacgacc	720
cccgcccatt gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc	780
attgacgtca atgggtggag tatttacggt aaactgccc cttggcagta catcaagtgt	840
atcatatgcc aagtccgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt	900
atgcccagta catgacctta cgggactttc ctacttggca gtacatctac gtattagtca	960
tcgctattac catggtgatg cggttttggc agtacaccaa tgggcgtgga tagcggtttg	1020
actcacgggg atttccaagt ctccaccca ttgacgtcaa tgggagtttg ttttggcacc	1080
aaaatcaacg ggactttcca aaatgtcgta ataacccgc cccgttgacg caaatgggcg	1140
gtaggcgtgt acggtgggag gtctatataa gcagagctcg tttagtgaac cgtcagatcg	1200

cctggaacgc	catccacgct	gttttgacct	ccatagaaga	cagcgatcgc	gccaccatgg	1260
ccgggatgtt	ccaggcactg	tccgaaggct	gcacacccta	tgatattaac	cagatgctga	1320
atgtcctggg	agaccaccag	gtctctggcc	tggagcagct	ggagagcatc	atcaacttcg	1380
agaagctgac	cgagtggaca	agctccaatg	tgatgcctat	cctgtcccca	ctgaccaagg	1440
gcatcctggg	cttcgtgttt	accctgacag	tgctttctga	gcggggcctg	tcttgcatca	1500
gcgaggcaga	cgcaaccaca	ccagagtccg	ccaatctggg	cgaggagatc	ctgtctcagc	1560
tgtacctgtg	gccccgggtg	acatatcact	ccccttctta	cgcttatcac	cagttcgagc	1620
ggagagccaa	gtacaagaga	cacttcccag	gctttggcca	gtctctgctg	ttcggctacc	1680
ccgtgtacgt	gttcggcgat	tgcgtgcagg	gcgactggga	tgccatccgg	tttagatact	1740
gcgcaccacc	tggatatgca	ctgctgaggt	gtaacgacac	caattattcc	gccctgctgg	1800
cagtgggcgc	cctggagggc	cctcgcaatc	aggattggct	gggcgtgcca	aggcagctgg	1860
tgacacgcat	gcaggccatc	cagaacgcag	gcctgtgcac	cctggtggca	atgctggagg	1920
agacaatctt	ctggctgcag	gcctttctga	tggccctgac	cgacagcggc	cccaagacaa	1980
acatcatcgt	ggattcccag	tacgtgatgg	gcatctccaa	gccttctttc	caggagtttg	2040
tggactggga	gaacgtgagc	ccagagctga	attccaccga	tcagccattc	tggcaggcag	2100
gaatcctggc	aaggaacctg	gtgcctatgg	tggccacagt	gcagggccag	aatctgaagt	2160
accagggcca	gagcctggtc	atcagcgcct	ccatcatcgt	gtttaacctg	ctggagctgg	2220
agggcgacta	tcgggacgat	ggcaacgtgt	gggtgcacac	cccactgagc	cccagaacac	2280
tgaacgcctg	ggtgaaggcc	gtggaggaga	agaagggcat	cccagtgcac	ctggagctgg	2340
cctccatgac	caatatggag	ctgatgtcta	gcatcgtgca	ccagcaggctg	aggacatacg	2400
gacccgtgtt	catgtgcctg	ggaggcctgc	tgaccatggg	ggcaggagcc	gtgtggctga	2460
cagtgcgggt	gctggagctg	ttcagagccg	cccagctggc	caacgatgtg	gtgctgcaga	2520
tcatggagct	gtgcggagca	gcctttcgcc	aggtgtgcca	caccacagtg	ccatggccca	2580
atgcctccct	gacccccaag	tggaacaatg	agacaacaca	gcctcagatc	gccaactgta	2640
gcgtgtacga	cttcttcgtg	tggctgcact	actatagcgt	gagggatacc	ctgtggcccc	2700
gcgtgacata	ccacatgaat	aagtacgcct	atcacatgct	ggagaggcgc	gccaagtata	2760
agagaggccc	tggcccaggc	gcaaagtttg	tggcagcatg	gaccctgaag	gccgccgccg	2820

gccccggccc	cggccagtat	atcaaggcta	acagtaagtt	cattggaatc	acagagctgg	2880
gacccggacc	tggataatga	gtttaaactc	ccattttaat	gtgagggtta	atgcttcgag	2940
cagacatgat	aagatacatt	gatgagtttg	gacaaaccac	aactagaatg	cagtgaaaaa	3000
aatgctttat	ttgtgaaatt	tgtgatgcta	ttgctttatt	tgtaacatt	ataagctgca	3060
ataaacaagt	taacaacaac	aattgcattc	attttatgtt	tcaggttcag	ggggagatgt	3120
gggagggtttt	ttaaagcaag	taaaacctct	acaaatgtgg	taaaataact	ataacgggtcc	3180
taaggtagcg	agtgagtagt	gttctggggc	gggggaggac	ctgcatgagg	gccagaataa	3240
ctgaaatctg	tgcttttctg	tgtgttgagc	cagcatgagc	ggaagcggct	cctttgaggg	3300
aggggtattc	agcccttatt	tgacggggcg	tctccctcc	tgggcgggag	tgcgtcagaa	3360
tgtgatggga	tccacggtgg	acggccggcc	cgtgcagccc	gcgaactctt	caaccctgac	3420
ctatgcaacc	ctgagctctt	cgtcgttgga	cgcagctgcc	gccgcagctg	ctgcatctgc	3480
cgccagcgcc	gtgcgcggaa	tggccatggg	cgccggctac	tacggcactc	tgggtggccaa	3540
ctcgagttcc	accaataatc	ccgccagcct	gaacgaggag	aagctgttgc	tgctgatggc	3600
ccagctcgag	gccttgaccc	agcgcctggg	cagctgacc	cagcagggtgg	ctcagctgca	3660
ggagcagacg	cgggccgcgg	ttgccacggt	gaaatccaaa	taaaaaatga	atcaataaat	3720
aaacggagac	ggttggtgat	tttaacacag	agtctgaatc	tttatgtgat	ttttcgcgcg	3780
cggtagggccc	tggaccaccg	gtctcgatca	ttgagcacc	ggtggatctt	ttccaggacc	3840
cggtagaggt	gggcttgat	gttgaggtac	atgggcatga	gcccgtccc	ggggtggagg	3900
tagctccatt	gcagggcctc	gtgctcgggg	gtgggtgttg	aaatcaccca	gtcatagcag	3960
gggcgcaggg	catggtgttg	cacaatatct	ttgaggagga	gactgatggc	cacgggcagc	4020
cctttggtgt	aggtgtttac	aaatctgttg	agctgggagg	gatgcatgcg	gggggagatg	4080
aggtgcatct	tggcctggat	cttgagattg	gcgatgttac	cggccagatc	ccgcctgggg	4140
ttcatgttgt	gcaggaccac	cagcacggtg	tatccggtgc	acttggggaa	tttatcatgc	4200
aacttggaag	ggaaggcgtg	aaagaatttg	gcgacgcctt	tgtgcccgcc	caggttttcc	4260
atgcactcat	ccatgatgat	ggcgatgggc	ccgtgggcgg	cggcctgggc	aaagacgttt	4320
cgggggtcgg	acacatcata	gttgtggtcc	tgggtgaggt	catcataggc	cattttaatg	4380
aatttggggc	ggagggtgcc	ggactggggg	acaaaggtag	cctcgatccc	gggggcgtag	4440

ttccccctcac agatctgcat ctcccaggct ttgagctcgg aggggggggat catgtccacc	4500
tgcggggcgga taaagaacac ggttttccggg gcggggggaga tgagctgggc cgaaagcaag	4560
ttccggagca gctgggactt gccgcagccg gtggggccgt agatgacccc gatgaccggc	4620
tgcaggtggt agttgaggga gagacagctg ccgtcctccc ggaggagggg ggccacctcg	4680
ttcatcatct cgcgcacgtg catgtttctg cgcaccagtt ccgccaggag gcgctctccc	4740
cccagggata ggagctcctg gagcgaggcg aagtttttca gcggcttgag tccgtcggcc	4800
atgggcattt tggagagggt ttgttgcaag agttccaggc ggtcccagag ctcggtgatg	4860
tgctctacgg catctcgatc cagcagacct cctcgtttcg cgggttgggg cggctgcggg	4920
agtagggcac cagacgatgg gcgtccagcg cagccagggt ccggtccttc cagggtcgca	4980
gcgtccgcgt cagggtggtc tccgtcacgg tgaaggggtg cgcgccgggc tgggcgcttg	5040
cgagggtgcg cttcaggctc atccggctgg tcgaaaaccg ctcccgatcg gcgccctgcg	5100
cgtcggccag gtagcaattg accatgagtt cgtagttgag cgcctcggcc gcgtggcctt	5160
tggcgcgagg cttacctttg gaagtctgcc cgcaggcggg acagaggagg gacttgaggg	5220
cgtagagctt gggggcgagg aagacggact cgggggcgta ggcgtccgcg ccgcagtggg	5280
cgcagacggt ctgcactcc acgagccagg tgaggtcggg ctggtcgggg tcaaaaacca	5340
gtttcccgcc gttctttttg atgcgtttct taccttttgt ctccatgagc tcgtgtcccc	5400
gctgggtgac aaagaggctg tccgtgtccc cgtagaccga ctttatgggc cggtcctcga	5460
gcggtgtgcc gcggtcctcc tcgtagagga accccgcccc ctccgagacg aaagcccggg	5520
tccaggccag cacgaaggag gccacgtggg acgggtagcg gtcgttgtcc accagcgggt	5580
ccaccttttc cagggtatgc aaacacatgt cccctcgtc cacatccagg aaggtgattg	5640
gcttgtaagt gtaggccacg tgaccggggg tcccggccgg gggggtataa aagggtgcgg	5700
gtccctgctc gtcctcactg tcttccggat cgctgtccag gagcgccagc tgttggggta	5760
ggtattccct ctcgaaggcg ggcatgacct cggcactcag gttgtcagtt tctagaaacg	5820
aggaggattt gatattgacg gtgccggcgg agatgccttt caagagcccc tcgtccatct	5880
ggtcagaaaa gacgatcttt ttgttgtcga gcttggtggc gaaggagccg tagagggcgt	5940
tggagaggag cttggcgatg gagcgcatgg tctggttttt ttccttgtcg gcgcgctcct	6000
tggcggcgat gttgagctgc acgtactcgc gcgccacgca cttccattcg gggaagacgg	6060

tggtcagctc gtcgggcacg attctgacct gccagccccg attatgcagg gtgatgaggt	6120
ccacactggg gccacctcg ccgcgccagg gctcattagt ccagcagagg cgtccgccct	6180
tgcgcgagca gaaggggggc aggggggtcca gcatgacctc gtcggggggg tcggcatcga	6240
tggtgaagat gccgggcagg aggtcgggggt caaagtagct gatggaagtg gccagatcgt	6300
ccagggcagc ttgccattcg cgcacggcca gcgcgcgctc gtagggactg aggggcgtgc	6360
cccagggcat gggatgggta agcgcggagg cgtacatgcc gcagatgtcg tagacgtaga	6420
ggggctcctc gaggatgccg atgtaggtgg ggtagcagcg ccccccgcgg atgctggcgc	6480
gcacgtagtc atacagctcg tgcgaggggg cgaggagccc cggggccagg ttggtgcgac	6540
tgggcctttc ggcgcggtag acgatctggc ggaaaatggc atgcgagttg gaggagatgg	6600
tgggcctttg gaagatgttg aagtgggcgt ggggcagtcc gaccgagtcg cggatgaagt	6660
gggcgtagga gtcttgacgc ttggcgacga gctcggcggg gactaggacg tccagagcgc	6720
agtagtcgag ggtctcctgg atgatgtcat acttgagctg tcccttttgt ttccacagct	6780
cgcggttgag aaggaactct tcgcggtcct tccagtactc ttcgaggggg aaccctcct	6840
gatctgcacg gtaagagcct agcatgtaga actggttgac ggcctttagt gcgcagcagc	6900
ccttctccac ggggagggcg taggcctggg cggccttgcg cagggaggtg tgcgtgaggg	6960
cgaaagtgtc cctgaccatg accttgagga actggtgctt gaagtcgata tcgtcgcagc	7020
ccccctgctc ccagagctgg aagtccgtgc gcttcttgta ggcggggttg ggcaaagcga	7080
aagtaacatc gttgaagagg atcttgcccg cgcggggcat aaagttgcga gtgatgcgga	7140
aaggttgggg cacctcggcc cggttgttga tgacctgggc ggcgagcacg atctcgtcga	7200
agccgttgat gttgtggccc acgatgtaga gttccacgaa tcgcggacgg cccttgacgt	7260
ggggcagttt cttgagctcc tcgtaggtga gtcgtcggg gtcgctgagc ccgtgctgct	7320
cgagcgccca gtcggcgaga tgggggttg cgcggaggaa ggaagtccag agatccacgg	7380
ccagggcggg ttgcagacgg tcccgtact gacggaactg ctgcccagc gccatTTTTT	7440
cgggggtgac gcagtagaag gtgcgggggt ccccggtcca gcgatcccat ttgagctgga	7500
gggcgagatc gagggcgagc tcgacgagcc ggtcgtcccc ggagagtttc atgaccagca	7560
tgaaggggac gagctgcttg ccgaaggacc ccatccaggt gtaggtttcc acatcgtagg	7620
tgaggaagag ctttctggtg cgaggatgcg agccgatggg gaagaactgg atctcctgcc	7680

accaattgga ggaatggctg ttgatgtgat ggaagtagaa atgccgacgg cgcgccgaac	7740
actcgtgctt gtgtttatac aagcggccac agtgctcgca acgctgcacg ggatgcacgt	7800
gctgcacgag ctgtacctga gttcctttga cgaggaatth cagtgggaag tggagtcgtg	7860
gcgctgcat ctcgtgctgt actacgtcgt ggtggtcggc ctggccctct tctgcctcga	7920
tggtggcat gctgacgagc ccgcgcggga ggcaggcca gacctcggcg cgagcgggtc	7980
ggagagcgag gacgagggcg cgcaggccgg agctgtccag ggtcctgaga cgctgcggag	8040
tcaggtcagt gggcagcggc ggcgcgcggt tgacttgca gagtttttcc agggcgcgcg	8100
ggaggtccag atggtacttg atctccaccg cgccattggt ggcgacgtcg atggcttgca	8160
gggtcccgtg cccctggggt gtgaccaccg tccccgttt cttcttgggc ggctggggcg	8220
acggggggcg tgcctcttcc atggttagaa gcggcggcga ggacgcgcgc cgggcggcag	8280
gggcggctcg gggcccggag gcaggggcgg caggggcacg tcggcgccgc gcgcgggtag	8340
gttctggtac tgcgcccga gaagactggc gtgagcgacg acgcgacggt tgacgtcctg	8400
gatctgacgc ctctgggtga aggccacggg acccgtgagt ttgaacctga aagagagttc	8460
gacagaatca atctcggtat cgttgacggc ggcctgccgc aggatctctt gcacgtcgcc	8520
cgagttgtcc tggtaggcga tctcggtcat gaactgctcg atctcctcct cttgaaggtc	8580
tccgcggccg gcgcgctcca cgggtggccgc gaggtcggtg gagatgcggc ccatgagctg	8640
cgagaaggcg ttcattgccc cctcgttcca gacgcggctg tagaccacga cgccctcggg	8700
atgcgggcg cgcatgacca cctgggcgag gttgagctcc acgtggcgcg tgaagaccgc	8760
gtagttgcag aggcgctggt agaggtagtt gagcgtgggt gcgatgtgct cggtagcgaa	8820
gaaatacatg atccagcggc ggagcggcat ctcgctgacg tcgcccagcg cctccaaacg	8880
ttccatggcc tcgtaaaagt ccacggcgaa gttgaaaaac tgggagttgc gcgccgagac	8940
ggtcaactcc tcctccagaa gacggatgag ctcggcgatg gtggcgcgca cctcgcgtc	9000
gaaggcccc gggagttcct ccacttcctc ttcttcctcc tccactaaca tctcttctac	9060
ttctcctca ggcggcagtg gtggcggggg agggggcctg cgtcgccggc ggcgcacggg	9120
cagacggtcg atgaagcgct cgatggtctc gccgcgccgg cgtcgcatgg tctcggtgac	9180
ggcgcgccc tcctcgcggg gccgcagcgt gaagacgcc cgcgcacatc ccagggtggc	9240
gggggggtcc ccgttgggca gggagagggc gctgacgatg catcttatca attgccccgt	9300

agggactccg cgcaaggacc tgagcgtctc gagatccacg ggatctgaaa accgctgaac	9360
gaaggcttcg agccagtcgc agtcgcaagg taggctgagc acggtttctt ctggcgggtc	9420
atgttggttg ggagcggggc gggcgatgct gctggtgatg aagttgaaat aggcggttct	9480
gagacggcgg atggtggcga ggagcaccag gtctttgggc ccggcttgct ggatgcgcag	9540
acggtcggcc atgccccagg cgtggtcctg acacctggcc aggtccttgt agtagtcctg	9600
catgagccgc tccacgggca cctcctcctc gcccgcgcg cgtgcatgc gcgtgagccc	9660
gaagccgcgc tggggctgga cgagcgccag gtcggcgacg acgcgctcgg cgaggatggc	9720
ttgctggatc tgggtgaggg tggctctgga gtcatcaaag tcgacgaagc ggtggtaggc	9780
tccggtgttg atggtgtagg agcagttggc catgacggac cagttgacgg tctggtggcc	9840
cggacgcacg agctcgtggt acttgaggcg cgagtaggcg cgcgtgtcga agatgtagtc	9900
gttgacagtg cgcaccaggt actggtagcc gatgaggaag tgcggcggcg gctggcggtg	9960
gagcggccat cgctcgttg cgggggcgcc gggcgcgagg tcctcgagca tggcgcggtg	10020
gtagccgtag atgtacctgg acatccaggt gatgccggcg gcggtggtgg aggcgcgcgg	10080
gaactcgcgg acgcggttcc agatgttcgc cagcggcagg aagtagttca tggtagggcac	10140
ggtctggccc gtgaggcgcg cgcagtcgtg gatgctctat acgggcaaaa acgaaagcgg	10200
tcagcggctc gactccgtgg cctggaggct aagcgaacgg gttgggctgc gcgtgtaccc	10260
cggttcgaat ctcgaatcag gctggagccg cagctaactg ggtattggca ctcccgtctc	10320
gaccaagcc tgcaccaacc ctccaggata cggaggcggg tcgttttgca actttttttt	10380
ggaggccgga tgagactagt aagcgcggaa agcggccgac cgcgatggct cgctgccgta	10440
gtctggagaa gaatcgccag ggttgcggtt cgggtgtgcc cggttcgagg ccggccggat	10500
tccgcggcta acgagggcgt ggctgccccg tcgtttccaa gaccccatag ccagccgact	10560
tctccagtta cggagcgagc ccctcttttg ttttgtttgt ttttgccaga tgcattcccgt	10620
actgcggcag atgcgcccc accaccctcc accgcaacaa cagccccctc cacagccggc	10680
gcttctgccc ccgccccagc agcaacttcc agccacgacc gcccgggccg ccgtgagcgg	10740
ggctggacag agttatgatc accagctggc cttggaagag ggcgaggggc tggcgcgctt	10800
gggggcgtcg tcgccggagc ggcacccgcg cgtgcagatg aaaagggacg ctcgcgaggc	10860
ctacgtgccc aagcagaacc tgttcagaga caggagcggc gaggagcccg aggagatgcg	10920

cgcgggcccgg	ttccacgcgg	ggcgggagct	gcggcgcggc	ctggaccgaa	agagggtgct	10980
gagggacgag	gatttcgagg	cggacgagct	gacggggatc	agccccgcgc	gcgcgcacgt	11040
ggccgcggcc	aacctggtca	cggcgtagca	gcagaccgtg	aaggaggaga	gcaacttcca	11100
aaaatccttc	aacaaccacg	tgcgcaccct	gatcgcgcg	gaggagggtga	ccctgggcct	11160
gatgcacctg	tgggacctgc	tggaggccat	cgtgcagaac	cccaccagca	agccgctgac	11220
ggcgcagctg	ttcctggtgg	tgcagcatag	tcgggacaac	gaagcgttca	gggaggcgct	11280
gctgaatatc	accgagcccc	agggccgctg	gctcctggac	ctggtgaaca	ttctgcagag	11340
catcgtggtg	caggagcgcg	ggctgccgct	gtccgagaag	ctggcggcca	tcaacttctc	11400
ggtgctgagt	ttgggcaagt	actacgctag	gaagatctac	aagaccccgt	acgtgccccat	11460
agacaaggag	gtgaagatcg	acgggtttta	catgcgcatg	accctgaaag	tgctgaccct	11520
gagcgacgat	ctgggggtgt	accgcaacga	caggatgcac	cgtgcggtga	gcgccagcag	11580
gcggcgcgag	ctgagcgacc	aggagctgat	gcatagtctg	cagcggggcc	tgaccggggc	11640
cgggaccgag	ggggagagct	actttgacat	gggcgcggac	ctgcactggc	agcccagccg	11700
ccgggccttg	gaggcggcgg	caggacccta	cgtagaagag	gtggacgatg	aggtggacga	11760
ggagggcgag	tacctggaag	actgatggcg	cgaccgtatt	tttgctagat	gcaacaacaa	11820
cagccacctc	ctgatcccg	gatgcgggcg	gcgctgcaga	gccagccgtc	cggcattaac	11880
tcctcggacg	attggaccca	ggccatgcaa	cgcatcatgg	cgctgacgac	ccgcaacccc	11940
gaagccttta	gacagcagcc	ccaggccaac	cggctctcgg	ccatcctgga	ggccgtggtg	12000
ccctcgcgct	ccaacccac	gcacgagaag	gtcctggcca	tcgtgaacgc	gctggtggag	12060
aacaaggcca	tccgcggcga	cgaggccggc	ctggtgtaca	acgcgctgct	ggagcgcgtg	12120
gcccgtaca	acagcaccaa	cgtgcagacc	aacctggacc	gcatggtgac	cgacgtgcgc	12180
gaggccgtgg	cccagcgcg	gcggttccac	cgcgagtcca	acctgggatc	catggtggcg	12240
ctgaacgcct	tcctcagcac	ccagcccggc	aacgtgcccc	ggggccagga	ggactacacc	12300
aacttcatca	gcgccctg	cctgatggtg	accgaggtgc	cccagagcga	ggtgtaccag	12360
tccgggcccgg	actacttctt	ccagaccagt	cgccagggct	tgcagaccgt	gaacctgagc	12420
caggctttca	agaacttgca	gggcctgtgg	ggcgtgcagg	ccccggtcgg	ggaccgcgcg	12480
acggtgtcga	gcctgctgac	gccgaactcg	cgcctgctgc	tgctgctggt	ggcccccttc	12540

acggacagcg	gcagcatcaa	ccgcaactcg	tacctgggct	acctgattaa	cctgtaccgc	12600
gaggccatcg	gccaggcgca	cgtggacgag	cagacctacc	aggagatcac	ccacgtgagc	12660
cgcgccctgg	gccaggacga	cccgggcaac	ctggaagcca	ccctgaactt	tttgctgacc	12720
aaccggtcgc	agaagatccc	gccccagtac	gcgctcagca	ccgaggagga	gcgcatcctg	12780
cgttacgtgc	agcagagcgt	gggcctgttc	ctgatgcagg	agggggccac	ccccagcgcc	12840
gcgctcgaca	tgaccgcgcg	caacatggag	cccagcatgt	acgccagcaa	ccgcccgttc	12900
atcaataaac	tgatggacta	cttgcatcgg	gcggccgcca	tgaactctga	ctatttcacc	12960
aacgccatcc	tgaatcccca	ctggctcccg	ccgccgggggt	tctacacggg	cgagtacgac	13020
atgcccgacc	ccaatgacgg	gttcctgttg	gacgatgtgg	acagcagcgt	gttctcccc	13080
cgaccgggtg	ctaacgagcg	ccccttgttg	aagaaggaag	gcagcgaccg	acgcccgtcc	13140
tcggcgctgt	ccggccgcga	gggtgctgcc	gcggcggtgc	ccgaggccgc	cagtcctttc	13200
ccgagcttgc	ccttctcgct	gaacagtatc	cgcagcagcg	agctgggcag	gatcacgcgc	13260
ccgcgcttgc	tgggcgaaga	ggagtacttg	aatgactcgc	tgttgagacc	cgagcgggag	13320
aagaacttcc	ccaataacgg	gatagaaagc	ctggtggaca	agatgagccg	ctggaagacg	13380
tatgcgcagg	agcacaggga	cgatccccgg	gcgtcgcagg	gggccacgag	ccggggcagc	13440
gccgcccgtg	aacgccggtg	gcacgacagg	cagcggggac	agatgtggga	cgatgaggac	13500
tccgccgacg	acagcagcgt	gttggaacttg	ggtgggagtg	gtaaccggtt	cgctcacctg	13560
cgcccccgta	tcgggcgcat	gatgtaagag	aaaccgaaaa	taaatgatac	tcaccaaggc	13620
catggcgacc	agcgtgcgtt	cgtttcttct	ctgttgttgt	tgtatctagt	atgatgaggc	13680
gtgcgtaccc	ggagggtcct	cctccctcgt	acgagagcgt	gatgcagcag	gcgatggcgg	13740
cggcggcgat	gcagcccccg	ctggaggctc	cttacgtgcc	cccgcggtac	ctggcgccta	13800
cggaggggcg	gaacagcatt	cgttactcgg	agctggcacc	cttgtagcat	accacccggt	13860
tgtacctggt	ggacaacaag	tcggcggaca	tcgcctcgct	gaactaccag	aacgaccaca	13920
gcaacttcct	gaccaccgtg	gtgcagaaca	atgacttcac	ccccacggag	gccagcacc	13980
agaccatcaa	ctttgacgag	cgctcgcggt	ggggcggcca	gctgaaaacc	atcatgcaca	14040
ccaacatgcc	caacgtgaac	gagttcatgt	acagcaacaa	gttcaaggcg	cgggtgatgg	14100
tctcccgcaa	gacccccaat	ggggtgacag	tgacagagga	ttatgatggt	agtcaggatg	14160

agctgaagta tgaatgggtg gaatttgagc tgcccgaagg caacttctcg gtgaccatga	14220
ccatcgacct gatgaacaac gccatcatcg acaattactt ggcggtgggg cggcagaacg	14280
gggtgctgga gagcgacatc ggcggtgaagt tcgacactag gaacttcagg ctgggctggg	14340
accccgtgac cgagctggtc atgcccgggg tgtacaccaa cgaggctttc catccccgata	14400
ttgtcttgct gcccggctgc ggggtggact tcaccgagag ccgcctcagc aacctgctgg	14460
gcattcgcaa gaggcagccc ttccaggaag gcttcagat catgtacgag gatctggagg	14520
ggggcaacat ccccgcgctc ctggatgtcg acgcctatga gaaaagcaag gaggatgcag	14580
cagctgaagc aactgcagcc gtagctaccg cctctaccga ggtcaggggc gataattttg	14640
caagcgccgc agcagtggca gcggccgagg cggctgaaac cgaaagtaag atagtcatc	14700
agccggtgga gaaggatagc aagaacagga gctacaacgt actaccggac aagataaaca	14760
ccgcctaccg cagctggtac ctagcctaca actatggcga ccccgagaag ggcgtagcgt	14820
cctggacgct gctcaccacc tcggacgtca cctgcggcgt ggagcaagtc tactggtcgc	14880
tgcccgacat gatgcaagac ccggtcacct tccgctccac gcgtcaagtt agcaactacc	14940
cggtaggtggg cgccgagctc ctgcccgtct actccaagag cttcttcaac gagcaggccg	15000
tctactcgca gcagctgcgc gccttcacct cgcttacgca cgtcttcaac cgcttccccg	15060
agaaccagat cctcgtccgc ccgcccgcgc ccaccattac caccgtcagt gaaaacgttc	15120
ctgctctcac agatcacggg accctgccgc tgcgcagcag tatccgggga gtccagcgcg	15180
tgaccgttac tgacgccaga cgccgcacct gccctacgt ctacaaggcc ctgggcatag	15240
tcgcgccgcg cgtcctctcg agccgcacct tctaaatgtc cattctcatc tcgccagta	15300
ataacaccgg ttggggcctg cgcgcgccca gcaagatgta cggaggcgct cgccaacgct	15360
ccacgcaaca ccccgtagcgc gtgcgcgggc acttccgcgc tccctggggc gccctcaagg	15420
gccgcgtgcg gtcgcgcacc accgtcgacg acgtgatcga ccaggtagtg gccgacgcgc	15480
gcaactacac ccccgccgcc gcgcccgtct ccaccgtgga cgccgtcatc gacagcgtgg	15540
tggccgacgc gcgccgtac gcccgccca agagccggcg gcggcgcatc gccggcgggc	15600
accggagcac ccccgccatg cgcgcggcgc gagccttgct gcgcagggcc aggcgcacgg	15660
gacgcagggc catgctcagg gcggccagac gcgcggcttc aggcgccagc gccggcagga	15720
cccgagacg cgcggccacg gcggcggcag cggccatcgc cagcatgtcc cgcccgcggc	15780

gagggAACgt	gtactgggtg	cgcgacgccg	ccaccggtgt	gcgcgtgccc	gtgcgcaccc	15840
gccccctcg	cacttgaaga	tgttcacttc	gcgatgttga	tgtgtcccag	cggcgaggag	15900
gatgtccaag	cgcaaattca	aggaagagat	gctccagggtc	atcgcgccctg	agatctacgg	15960
ccctgcggtg	gtgaaggagg	aaagaaagcc	ccgcaaaatc	aagcgggtca	aaaaggacaa	16020
aaaggaagaa	gaaagtgatg	tggacggatt	ggtggagttt	gtgcgcgagt	tcgccccccg	16080
gcggcgcggtg	cagtggcgcg	ggcgggaaggt	gcaaccggtg	ctgagacccg	gcaccaccgt	16140
ggtcttcacg	cccggcgagc	gctccggcac	cgcttccaag	cgctcctacg	acgagggtgta	16200
cggggatgat	gatattctgg	agcaggcggc	cgagcgccctg	ggcgagtttg	cttacggcaa	16260
gcgcagccgt	tccgcaccga	aggaagaggc	ggtgtccatc	ccgtggacc	acggcaaccc	16320
cacgccgagc	ctcaagcccg	tgaccttgca	gcaggtgctg	ccgaccgcgg	cgccgcgccg	16380
ggggttcaag	cgcgagggcg	aggatctgta	ccccaccatg	cagctgatgg	tgcccaagcg	16440
ccagaagctg	gaagacgtgc	tggagaccat	gaaggtggac	ccggacgtgc	agccccgaggt	16500
caaggtgcgg	cccatcaagc	aggtggcccc	gggcctgggc	gtgcagaccg	tggacatcaa	16560
gattcccacg	gagcccatgg	aaacgcagac	cgagcccatg	atcaagccca	gcaccagcac	16620
catggagggtg	cagacggatc	cctggatgcc	atcggctcct	agtcgaagac	cccggcgcaa	16680
gtacggcgcg	gccagcctgc	tgatgccccaa	ctacgcgctg	catccttcca	tcatccccac	16740
gccgggctac	cgcggcacgc	gcttctaccg	cggtcatacc	agcagccgcc	gccgcaagac	16800
caccactcgc	cgccgccgtc	gccgcaccgc	cgctgcaacc	acccttgccg	ccctgggtgcg	16860
gagagtgtac	cgccgcggcc	gcgcacctct	gaccctgccg	cgcgcgcgct	accacccgag	16920
catcgccatt	taaactttcg	cctgctttgc	agatcaatgg	ccctcacatg	ccgccttcgc	16980
gttcccatta	cgggctaccg	aggaagaaaa	ccgcgccgta	gaaggctggc	ggggaacggg	17040
atgcgtcgcc	accaccaccg	gcggcgggcg	gccatcagca	agcggttggg	gggaggcttc	17100
ctgcccgcgc	tgatccccat	catcgccgcg	gcgatcgggg	cgatccccgg	cattgcttcc	17160
gtggcggtgc	aggcctctca	gcgccactga	gacacacttg	gaaacatctt	gtaataaacc	17220
aatggactct	gacgctcctg	gtcctgtgat	gtgttttcgt	agacagatgg	aagacatcaa	17280
tttttcgtcc	ctggctccgc	gacacggcac	gcggccgttc	atgggcacct	ggagcgacat	17340
cggcaccagc	caactgaacg	ggggcgccctt	caattggagc	agtctctgga	gcgggcttaa	17400

gaatttcggg	tccacgctta	aaacctatgg	cagcaaggcg	tggaacagca	ccacagggca	17460
ggcgctgagg	gataagctga	aagagcagaa	cttccagcag	aaggtggtcg	atgggctcgc	17520
ctcgggcatc	aacggggtgg	tggacctggc	caaccaggcc	gtgcagcggc	agatcaacag	17580
ccgcctggac	ccggtgccgc	ccgccggctc	cgtggagatg	ccgcagggtg	aggaggagct	17640
gcctccccctg	gacaagcggg	gcgagaagcg	accccgcgcc	gatgcggagg	agacgctgct	17700
gacgcacacg	gacgagccgc	ccccgtacga	ggaggcgggtg	aaactgggtc	tgcccaccac	17760
gcggcccatc	gcgcccctgg	ccaccgggggt	gctgaaaccc	gaaaagcccg	cgaccctgga	17820
cttgccctcct	ccccagcctt	ccgccccctc	tacagtggct	aagcccctgc	cgccggtggc	17880
cgtggcccg	gcgcgacccg	ggggcaccgc	ccgccctcat	gcgaactggc	agagcactct	17940
gaacagcatc	gtgggtctgg	gagtgcagag	tgtgaagcgc	cgccgctgct	attaaaccta	18000
ccgtagcgct	taacttgctt	gtctgtgtgt	gtatgtatta	tgtcgccgcc	gccgctgtcc	18060
accagaagga	ggagtgaaga	ggcgcgtcgc	cgagttgcaa	gatggccacc	ccatcgatgc	18120
tgccccagtg	ggcgtacatg	cacatcgccg	gacaggacgc	ttcggagtac	ctgagtccgg	18180
gtctggtgca	gtttgcccgc	gccacagaca	cctacttcag	tctggggaac	aagttagga	18240
acccacgggt	ggcgcccacg	cacgatgtga	ccaccgaccg	cagccagcgg	ctgacgctgc	18300
gcttcgtgcc	cgtggaccgc	gaggacaaca	cctactcgta	caaagtgcgc	tacacgctgg	18360
ccgtgggcga	caaccgcgtg	ctggacatgg	ccagcaccta	ctttgacatc	cgcggcgtgc	18420
tggatcgggg	ccctagcttc	aaaccctact	ccggcaccgc	ctacaacagt	ctggcccca	18480
agggagcacc	caacacttgt	cagtggacat	ataaagccga	tggtgaaact	gccacagaaa	18540
aaacctatac	atatggaaat	gcacccgtgc	agggcattaa	catcacaaaa	gatggtattc	18600
aacttggaac	tgacaccgat	gatcagccaa	tctacgcaga	taaaacctat	cagcctgaac	18660
ctcaagtggg	tgatgctgaa	tggcatgaca	tcactggtac	tgatgaaaag	tatggaggca	18720
gagctcttaa	gcctgatacc	aaaatgaagc	cttgttatgg	ttcttttgcc	aagcctacta	18780
ataaagaagg	aggtcaggca	aatgtgaaaa	caggaacagg	cactactaaa	gaatatgaca	18840
tagacatggc	tttctttgac	aacagaagtg	cggctgctgc	tggcctagct	ccagaaattg	18900
ttttgtatac	tgaaaatgtg	gattttgaaa	ctccagatac	ccatattgta	tacaaagcag	18960
gcacagatga	cagcagctct	tctattaatt	tgggtcagca	agccatgccc	aacagaccta	19020

actacattgg	tttcagagac	aacttttatcg	ggctcatgta	ctacaacagc	actggcaata	19080
tgggggtgct	ggccggtcag	gcttctcagc	tgaatgctgt	ggttgacttg	caagacagaa	19140
acaccgagct	gtcctaccag	ctcttgcttg	actctctggg	tgacagaacc	cggtatttca	19200
gtatgtggaa	tcaggcgggtg	gacagctatg	atcctgatgt	gcgcattatt	gaaaatcatg	19260
gtgtggagga	tgaacttccc	aactattgtt	tccctctgga	tgctgttggc	agaacagata	19320
cttatcaggg	aattaaggct	aatggaactg	atcaaaccac	atggaccaa	gatgacagt	19380
tcaatgatgc	taatgagata	ggcaagggtg	atccattcgc	catggaaatc	aacatccaag	19440
ccaacctgtg	gaggaacttc	ctctacgcca	acgtggccct	gtacctgccc	gactcttaca	19500
agtacacgcc	ggccaatgtt	accctgcccc	ccaacaccaa	cacctacgat	tacatgaacg	19560
gccgggtggg	ggcgccctcg	ctggtggact	cctacatcaa	catcggggcg	cgctggtcgc	19620
tggatcccat	ggacaacgtg	aacccttca	accaccaccg	caatgcgggg	ctgcgctacc	19680
gctccatgct	cctgggcaac	gggcgctacg	tgcccttcca	catccagggtg	ccccagaaat	19740
ttttcgccat	caagagcctc	ctgctcctgc	ccgggtccta	cacctacgag	tggaacttcc	19800
gcaaggacgt	caacatgatc	ctgcagagct	ccctcggcaa	cgacctgcgc	acggacgggg	19860
cctccatctc	cttcaccagc	atcaacctct	acgccacctt	cttccccatg	gcgcacaaca	19920
cggcctccac	gctcgaggcc	atgctgcgca	acgacaccaa	cgaccagtcc	ttcaacgact	19980
acctctcggc	ggccaacatg	ctctaccccc	tcccggccaa	cgccaccaac	gtgcccattct	20040
ccatcccctc	gcgcaactgg	gccgccttcc	gcggctgggtc	cttcacgcgt	ctcaagacca	20100
aggagacgcc	ctcgctgggc	tccgggttcg	accctactt	cgtctactcg	ggctccatcc	20160
cctacctcga	cggcaccttc	tacctcaacc	acaccttcaa	gaaggctctcc	atcaccttcg	20220
actcctccgt	cagctggccc	ggcaacgacc	ggctcctgac	gccaacgag	ttcgaaatca	20280
agcgcaccgt	cgacggcgag	ggctacaacg	tggcccagtg	caacatgacc	aaggactggt	20340
tcctgggtcca	gatgctggcc	cactacaaca	tcggctacca	gggcttctac	gtgcccagg	20400
gctacaagga	ccgcatgtac	tccttcttcc	gcaacttcca	gcccattgagc	cgccagggtg	20460
tggacgagggt	caactacaag	gactaccagg	ccgtcacctt	ggcctaccag	cacaacaact	20520
cgggcttcgt	cggctacctc	gcgcccacca	tgcgccagg	ccagccctac	cccgccaact	20580
acccctaccc	gctcatcggc	aagagcgccg	tcaccagcgt	caccagaaa	aagttcctct	20640

gcgacagggt catgtggcgc atcccccttct ccagcaactt catgtccatg ggcgcgctca	20700
ccgacctcgg ccagaacatg ctctatgcca actccgcccc cgcgctagac atgaatttcg	20760
aagtcgaccc catggatgag tccacccttc tctatgttgt cttcgaagtc ttcgacgtcg	20820
tccgagtga ccagccccac cgcgggcgtca tcgaggccgt ctacctgcgc acccccttct	20880
cggccggtaa cgccaccacc taagctcttg cttcttgcaa gccatggccg cgggctccgg	20940
cgagcaggag ctcagggccca tcatccgcga cctgggctgc gggccctact tcctgggcac	21000
cttcgataag cgcttcccgg gattcatggc cccgcacaag ctggcctgcg ccatcgtcaa	21060
cacggccggc cgcgagaccg ggggcgagca ctggctggcc ttcgcctgga acccgcgctc	21120
gaacacctgc tacctcttcg accccttcgg gttctcggac gagcgcctca agcagatcta	21180
ccagttcgag tacgagggcc tgctgcgccg cagcgccctg gccaccgagg accgctgcgt	21240
caccctggaa aagtccaccc agaccgtgca gggctccgcgc tcggccgcct gcgggctctt	21300
ctgctgcatg ttcttcgacg ccttcgtgca ctggcccgac cgcccatgg acaagaaccc	21360
caccatgaac ttgctgacgg ggggtgcccc cggcattgctc cagtcgcccc aggtggaacc	21420
caccctgcgc cgcaaccagg aggcgctcta ccgcttcctc aactcccact ccgcctactt	21480
tcgctccac cgcgcgcgca tcgagaaggc caccgccttc gaccgcatga atcaagacat	21540
gtaaaccgtg tgtgtatgtt aaatgtcttt aataaacagc actttcatgt tacacatgca	21600
tctgagatga tttatttaga aatcgaaagg gttctgccgg gtctcggcat ggcccgcggg	21660
caggacacg ttgcggaact ggtacttggc cagccacttg aactcgggga tcagcagttt	21720
gggcagcggg gtgtcgggga aggagtcggt ccacagcttc cgcgtcagtt gcaggcgcc	21780
cagcaggtcg ggcgcggaga tcttgaaatc gcagttggga cccgcgttct gcgcgcggga	21840
gttgcggtac acggggttgc agcactggaa caccatcagg gccgggtgct tcacgctcgc	21900
cagcaccgtc gcgtcggtaga tgctctccac gtcgaggtcc tcggcgttgg ccatcccgaa	21960
gggggtcatc ttgcaggtct gccttcccat ggtgggcacg caccgggct tgtggttgca	22020
atcgcagtg agggggatca gcatcatctg ggcctggtcg gcgttcatcc ccgggtacat	22080
ggccttcatg aaagcctcca attgcctgaa cgctgtctgg gccttggctc cctcggtgaa	22140
gaagaccccg caggacttgc tagagaactg gttggtggcg caccggcgt cgtgcacgca	22200
gcagcgcgcg tcgttggttg ccagctgcac cacgtgcgc cccagcgggt tctgggtgat	22260

cttggcccgg tcggggttct ccttcagcgc gcgctgcccg ttctcgctcg ccacatccat	22320
ctcgatcatg tgctccttct ggatcatggg ggtcccgtgc aggcaccgca gcttgccctc	22380
ggcctcgggtg caccctgtga gccacagcgc gcacccgggtg cactcccagt tcttgtgggc	22440
gatctgggaa tgcgcgtgca cgaagccctg caggaagcgg cccatcatgg tggtcagggt	22500
cttgttgcta gtgaaggta gcggaatgcc gcgggtgctc tcgttgatgt acagggtggca	22560
gatgcggcgg tacacctcgc cctgctcggg catcagctgg aagttggctt tcaggtcgggt	22620
ctccacgcgg tagcgggtcca tcagcatagt catgatattc ataccttct cccaggccga	22680
gacgatgggc aggctcatag ggttcttcac catcatctta gcgctagcag ccgcggccag	22740
ggggtcgctc tcgtccaggg tctcaaagct ccgcttgccg tccttctcgg tgatccgcac	22800
cggggggtag ctgaagccca cggccgccag ctctcctcg gcctgtcttt cgtcctcgct	22860
gtcctggctg acgtcctgca ggaccacatg cttgggtctg cggggtttct tcttgggcgg	22920
cagcggcggc ggagatgttg gagatggcga gggggagcgc gagttctcgc tcaccactac	22980
tatctcttcc tcttcttggg ccgaggccac gcggcggtag gtatgtctct tcgggggcag	23040
aggcggaggc gacgggctct cgccgccgcg acttggcgga tggctggcag agccccctcc	23100
gcgttcgggg gtgcgctccc ggcggcgctc tgactgactt cctccgcggc cggccattgt	23160
gttctcctag ggaggaacaa caagcatgga gactcagcca tcgccaacct cgccatctgc	23220
ccccaccgcc gacgagaagc agcagcagca gaatgaaagc ttaaccgccc cgccgccag	23280
ccccgccacc tccgacgcgg ccgtcccaga catgcaagag atggaggaat ccatcgagat	23340
tgacctgggc tatgtgacgc ccgcggagca cgaggaggag ctggcagtgc gcttttcaca	23400
agaagagata caccaagaac agccagagca ggaagcagag aatgagcaga gtcaggctgg	23460
gctcgagcat gacggcgact acctccacct gagcgggggg gaggacgcgc tcatcaagca	23520
tctggcccgg caggccacca tcgtcaagga tgcgctgctc gaccgcaccg aggtgcccct	23580
cagcgtggag gagctcagcc gcgcctacga gttgaacctc ttctcgccgc gcgtgcccc	23640
caagcgccag cccaatggca cctgcgagcc caaccgcgc ctcaacttct acccggtctt	23700
cgcggtgccc gaggccctgg ccacctacca catctttttc aagaacaaa agatccccgt	23760
ctcctgccgc gccaaaccga cccgcgccga cgcccttttc aacctgggtc ccggcgcccc	23820
cctacctgat atcgctctct tggaagaggt tcccaagatc ttcgagggtc tgggcagcga	23880

cgagactcgg gccgcgaacg ctctgcaagg agaaggagga gagcatgagc accacagcgc	23940
cctggtcgag ttggaaggcg acaacgcgcg gctggcggtg ctcaaacgca cggtcgagct	24000
gacccatttc gcctaccgag ctctgaacct gcccccaaa gtcatgagcg cggtcatgga	24060
ccaggtgctc atcaagcgcg cgtcgcccat ctccgaggac gagggcatgc aagactccga	24120
ggagggcaag cccgtggtca gcgacgagca gctggcccgg tggctgggtc ctaatgctag	24180
tccccagagt ttggaagagc ggcgcaaact catgatggcc gtggtcctgg tgaccgtgga	24240
gctggagtgc ctgcgccgct tcttcgccga cgcggagacc ctgcgcaagg tcgaggagaa	24300
cctgcactac ctcttcaggc acgggttcgt gcgccaggcc tgcaagatct ccaacgtgga	24360
gctgaccaac ctggtctcct acatgggcat cttgcacgag aaccgcctgg ggcagaacgt	24420
gctgcacacc accctgcgcg gggaggcccg gcgcgactac atccgcgact gcgtctacct	24480
ctacctctgc cacacctggc agacgggcat gggcgtgtgg cagcagtgtc tggaggagca	24540
gaacctgaaa gagctctgca agctcctgca gaagaacctc aagggtctgt ggaccgggtt	24600
cgacgagcgc accaccgcct cggacctggc cgacctcatt ttccccgagc gcctcaggct	24660
gacgctgcgc aacggcctgc ccgactttat gagccaaagc atgttgcaaa actttcgtc	24720
tttcatcctc gaacgctccg gaatcctgcc cgccacctgc tccgcgctgc cctcggactt	24780
cgtgccgctg accttccgcg agtgcccccc gccgctgtgg agccactgct acctgctgcg	24840
cctggccaac tacctggcct accactcgga cgtgatcgag gacgtcagcg gcgagggcct	24900
gctcgagtgc cactgccgct gcaacctctg cacgccgcac cgctccctgg cctgcaaccc	24960
ccagctgctg agcgagaccc agatcatcgg caccttcgag ttgcaagggc ccagcgaagg	25020
cgaggggttca gccgccaagg ggggtctgaa actcaccgcc gggctgtgga cctcggccta	25080
cttgcgcaag ttcgtgcccc aggactacca tcccttcgag atcaggttct acgaggacca	25140
atcccatccg cccaaggccg agctgtcggc ctgcgtcatc acccaggggg cgatcctggc	25200
ccaattgcaa gccatccaga aatcccgcc aagaattcttg ctgaaaaagg gccgcgggggt	25260
ctacctcgac cccagaccg gtgaggagct caaccccggc ttccccagg atgccccgag	25320
gaaacaagaa gctgaaagtg gagctgccgc ccgtggagga tttggaggaa gactgggaga	25380
acagcagtca ggcagaggag gaggagatgg aggaagactg ggacagcact caggcagagg	25440
aggacagcct gcaagacagt ctggaggaag acgaggagga ggcagaggag gaggtggaag	25500

aagcagccgc	cgccagaccg	tcgtcctcgg	cgggggagaa	agcaagcagc	acggatacca	25560
tctccgctcc	gggtcgggggt	cccgtcgcac	cacacagtag	atgggacgag	accggacgat	25620
tcccgaaccc	caccacccag	accggttaaga	aggagcggca	gggatacaag	tcctggcggg	25680
ggcacaaaaa	cgccatcgtc	tcctgcttgc	aggcctgcgg	gggcaacatc	tccttcaccc	25740
ggcgctacct	gctctttcac	cgcggggtga	actttccccg	caacatcttg	cattactacc	25800
gtcacctcca	cagcccctac	tactttcaag	aagaggcagc	agcagcagaa	aaagaccagc	25860
agaaaaccag	cagctagaaa	atccacagcg	gcggcagcag	gtggactgag	gatcgcgggc	25920
aacgagccgg	cgcaaaccgc	ggagctgagg	aaccggatct	ttcccaccct	ctatgccatc	25980
ttccagcaga	gtcgggggca	ggagcaggaa	ctgaaagtca	agaaccgttc	tctgcgctcg	26040
ctcaccgcga	gttgtctgta	tcacaagagc	gaagaccaac	ttcagcgcac	tctcgaggac	26100
gccgaggctc	tcttcaacaa	gtactgcgcg	ctcactctta	aagagtagcc	cgcgcccgcc	26160
cagtcgcaga	aaaaggcggg	aattacgtca	cctgtgccct	tcgccctagc	cgccctccacc	26220
catcatcatg	agcaaagaga	ttcccacgcc	ttacatgtgg	agctaccagc	cccagatggg	26280
cctggccgcc	ggtgccgccc	aggactactc	caccgcgatg	aattggctca	gcgccggggc	26340
cgcgatgatc	tcacgggtga	atgacatccg	cgcccaccga	aaccagatac	tcctagaaca	26400
gtcagcgctc	accgccacgc	cccgcaatca	cctcaatccg	cgtaattggc	ccgccgccct	26460
gggtgtaccag	gaaattcccc	agcccacgac	cgtactactt	ccgcgagacg	cccaggccga	26520
agtccagctg	actaactcag	gtgtccagct	ggcgggcggc	gccaccctgt	gtcgtcaccg	26580
ccccgctcag	ggtataaagc	ggctggtgat	ccggggcaga	ggcacacagc	tcaacgacga	26640
gggtggtgagc	tcttcgctgg	gtctgcgacc	tgacggagtc	ttccaactcg	ccggatcggg	26700
gagatcttcc	ttcacgcctc	gtcaggccgt	cctgactttg	gagagtctgt	cctcgcagcc	26760
ccgctcgggt	ggcatcggca	ctctccagtt	cgtggaggag	ttcactccct	cggtctactt	26820
caacccttc	tccggctccc	ccggccacta	cccggacgag	ttcatcccga	acttcgacgc	26880
catcagcgag	tcgggtggacg	gctacgattg	aaactaatca	cccccttatc	cagtgaataa	26940
aagatcatat	tgatgatgat	tttacagaaa	taaaaaataa	tcatttgatt	tgaaataaag	27000
atacaatcat	attgatgatt	tgagtttaac	aaaaaaataa	agaatcactt	acttgaaatc	27060
tgataccagg	tctctgtcca	tgttttctgc	caacaccact	tcactcccct	cttcccagct	27120

ctggtactgc	aggccccggc	gggctgcaaa	cttcctccac	acgctgaagg	ggatgtcaaa	27180
ttcctcctgt	ccctcaatct	tcattttatc	ttctatcaga	tgtccaaaaa	gcgcgtccgg	27240
gtggatgatg	acttcgaccc	cgtctacccc	tacgatgcag	acaacgcacc	gaccgtgccc	27300
ttcatcaacc	cccccttcgt	ctcttcagat	ggattccaag	agaagcccct	gggggtgttg	27360
tccctgcgac	tggccgaccc	cgtcaccacc	aagaacgggg	aatcaccct	caagctggga	27420
gaggggggtg	acctcgattc	ctcgggaaaa	ctcatctcca	acacggccac	caaggccgcc	27480
gcccctctca	gtttttccaa	caacaccatt	tcccttaaca	tggatcaccc	cttttacact	27540
aaagatggaa	aattatcctt	acaagtttct	ccaccattaa	atatactgag	aacaagcatt	27600
ctaaacacac	tagctttagg	ttttggatca	ggtttaggac	tccgtggctc	tgccttggca	27660
gtacagttag	tctctccact	tacatttgat	actgatggaa	acataaagct	taccttagac	27720
agaggtttgc	atgttacaac	aggagatgca	attgaaagca	acataagctg	ggctaaaggt	27780
ttaaaatttg	aagatggagc	catagcaacc	aacattggaa	atgggtaga	gtttggaagc	27840
agtagtacag	aaacaggtgt	tgatgatgct	tacccaatcc	aagttaaact	tggatctggc	27900
cttagctttg	acagtacagg	agccataatg	gctggtaaca	aagaagacga	taaactcact	27960
ttgtggacaa	cacctgatcc	atcaccaaac	tgtcaaatac	tcgcagaaaa	tgatgcaaaa	28020
ctaacacttt	gcttgactaa	atgtggtagt	caaatactgg	ccactgtgtc	agtcttagtt	28080
gtaggaagtg	gaaacctaaa	ccccattact	ggcaccgtaa	gcagtgtctc	ggtgtttcta	28140
cgttttgatg	caaacggtgt	tcttttaaca	gaacattcta	cactaaaaaa	atactggggg	28200
tataggcagg	gagatagcat	agatggcact	ccatatacca	atgctgtagg	attcatgccc	28260
aatttaaaag	cttatccaaa	gtcacaaagt	tctactacta	aaaataatat	agtagggcaa	28320
gtatacatga	atggagatgt	ttcaaaacct	atgcttctca	ctataaccct	caatggtact	28380
gatgacagca	acagtacata	ttcaatgtca	ttttcataca	cctggactaa	tggaagctat	28440
gttgagcaa	catttggggc	taactcttat	accttctcat	acatcgccca	agaatgaaca	28500
ctgtatccca	ccctgcatgc	caacccttcc	cacccactc	tgtggaacaa	actctgaaac	28560
acaaaataaa	ataaagttca	agtgttttat	tgattcaaca	gttttacagg	attcgagcag	28620
ttatttttcc	tccaccctcc	caggacatgg	aatacaccac	cctctcccc	cgcacagcct	28680
tgaacatctg	aatgccattg	gtgatggaca	tgcttttggg	ctccacgttc	cacacagttt	28740

cagagcgagc cagtctcggg tcggtcaggg agatgaaacc ctccgggcac tcccgcattct	28800
gcacctcaca gctcaacagc tgaggattgt cctcgggtggc cgggatcacg gttatctgga	28860
agaagcagaa gagcggcggc gggaatcata gtccgcgaac gggatcggcc ggtggtgtcg	28920
catcaggccc cgcagcagtc gctgccgccg ccgctccgtc aagctgctgc tcagggggtc	28980
cgggtccagg gactccctca gcatgatgcc cacggccctc agcatcagtc gtctggtgcg	29040
gcgggcgcag cagcgcattgc ggatctcgct caggctcgctg cagtacgtgc aacacagaac	29100
caccagggtt ttcaacagtc catagttcaa cacgctccag ccgaaactca tcgcgggaag	29160
gatgctaccc acgtggccgt cgtaccagat cctcaggtaa atcaagtggc gccccctcca	29220
gaacacgctg cccacgtaca tgatctcctt gggcatgtgg cggttcacca cctcccggta	29280
ccacatcacc ctctggttga acatgcagcc ccggatgatc ctgcggaacc acagggccag	29340
caccgccccg cccgccatgc agcgaagaga ccccggttcc cggcaatggc aatggaggac	29400
ccaccgctcg taccctgga tcattctgga gctgaacaag tctatgttgg cacagcacag	29460
gcatatgctc atgcatctct tcagcactct caactcctcg ggggtcaaaa ccatatccca	29520
gggcacgggg aactcttgca ggacagcga cccgcagaa cagggaatc ctcgcacaga	29580
acttacattg tgcattgaca gggatctgca atcaggcagc accgggtgat cctccaccag	29640
agaagcgagg gtctcggctc cctcacagcg tggtaagggg gccggccgat acgggtgatg	29700
gcgggacgag gctgatcgtg ttcgcgaccg tgtcatgatg cagttgcttt cggacatttt	29760
cgtacttgct gtagcagaac ctggtccggg cgctgcacac cgatcgccgg cggcggcttc	29820
ggcgcttgga acgctcggcg ttgaaattgt aaaacagcca ctctctcaga ccgtgcagca	29880
gatctagggc ctcaggagtg atgaagatcc catcatgcct gatggctctg atcacatcga	29940
ccaccgtgga atgggcccaga cccagccaga tgatgcaatt ttgttgggtt tcggtgacgg	30000
cgggggaggg aagaacagga agaaccatga ttaactttta atccaaacgg tctcggagta	30060
cttcaaaatg aagatcgagg agatggcacc tctcgcccc gctgtgttgg tggaaaataa	30120
cagccaggtc aaagggtgata cggttctcga gatgttcac ggtggcttcc agcaaagcct	30180
ccacgcgcac atccagaaac aagacaatag cgaaagcggg agggttctct aattcctcaa	30240
tcatcatgtt aactcctgc accatcccca gataattttc atttttccag ccttgaatga	30300
ttcgaactag ttcttgaggt aaatccaagc cagccatgat aaagagctcg cgcagagcgc	30360

cctccaccgg cattcttaag cacaccctca taattccaag atattctgct cctggttcac	30420
ctgcagcaga ttgacaagcg gaatatcaaa atctctgccg cgatccctga gtcctccct	30480
cagcaataac tgtaagtact ctttcatatc ctctccgaaa ttttttagcca taggaccacc	30540
aggaataaga ttagggcaag ccacagtaca gataaaccga agtcctcccc agtgagcatt	30600
gccaaatgca agactgctat aagcatgctg gctagacccg gtgatatctt ccagataact	30660
ggacagaaaa tcgcccaggc aatttttaag aaaatcaaca aaagaaaaat cctccaggtg	30720
gacgtttaga gcctcgggaa caacgatgaa gtaaattgcaa gcggtgcgtt ccagcatggt	30780
tagttagctg atctgtagaa aaaacaaaaa tgaacattaa accatgctag cctggcgaac	30840
aggtgggtaa atcgtttctt ccagcaccag gcaggccacg ggggtctccgg cgcgaccctc	30900
gtaaaaattg tcgctatgat tgaaaaccat cacagagaga cgttcccggg ggccggcgtg	30960
aatgattcga caagatgaat acacccccgg aacattggcg tccgcgagtg aaaaaagcg	31020
cccgaggaag caataaggca ctacaatgct cagtctcaag tccagcaaag cgatgccatg	31080
cggatgaagc acaaaattct caggtgcgta caaaatgtaa ttactcccct cctgcacagg	31140
cagcaaagcc ccgatccct ccaggtacac atacaaagcc tcagcgtcca tagcttaccg	31200
agcagcagca cacaacaggc gcaagagtca gagaaaggct gagctctaac ctgtccaccc	31260
gctctctgct caatatatag ccagatcta cactgacgta aaggccaaag tctaaaaata	31320
cccgccaaat aatcacacac gcccagcaca cgcccagaaa ccggtgacac actcaaaaaa	31380
atacgcgcac ttctcaaac gcccaaaact gccgtcattt ccgggttccc acgctacgtc	31440
atcaaaacac gactttcaaa ttccgtcgac cgttaaaaac gtcacccgcc ccgcccctaa	31500
cggtcgcccg tctctcagcc aatcagcgcc ccgcatcccc aaattcaaac acctcatttg	31560
catattaacg cgcacaaaaa gtttgagg	31588

<210> 3

<211> 11447

<212> DNA

<213> Venezuelan equine encephalitis virus

<400> 3

atgggcggcg catgagagaa gccagacca attacctacc caaatggag aaagttcacg	60

ttgacatcga ggaagacagc ccattcctca gagctttgca gcggagcttc ccgagtttg	120

aggtagaagc caagcaggtc actgataatg accatgctaa tgccagagcg ttttcgcatc	180
tggcttcaaa actgatcgaa acggaggtgg acccatccga cacgatcctt gacattggaa	240
gtgcgcccgc ccgcagaatg tattctaagc acaagtatca ttgtatctgt ccgatgagat	300
gtgcggaaga tccggacaga ttgtataagt atgcaactaa gctgaagaaa aactgtaagg	360
aaataactga taaggaattg gacaagaaaa tgaaggagct cgccgccgtc atgagcgacc	420
ctgacctgga aactgagact atgtgcctcc acgacgacga gtcgtgtcgc tacgaagggc	480
aagtcgctgt ttaccaggat gtatacgcgg ttgacggacc gacaagtctc tatcaccaag	540
ccaataaggg agttagagtc gcctactgga taggctttga caccaccctt tttatgttta	600
agaacttggc tggagcatat ccatcatact ctaccaactg ggccgacgaa accgtgttaa	660
cggctcgtaa cataggccta tgcagctctg acgttatgga gcggtcacgt agagggatgt	720
ccattcttag aaagaagtat ttgaaaccat ccaacaatgt tctattctct gttggctcga	780
ccatctacca cgagaagagg gacttactga ggagctggca cctgccgtct gtatttcact	840
tacgtggcaa gcaaaattac acatgtcggg gtgagactat agttagttgc gacgggtacg	900
tcgttaaaag aatagctatc agtccaggcc tgtatgggaa gccttcaggc tatgctgcta	960
cgatgcaccg cgagggattc ttgtgctgca aagtgacaga cacattgaac ggggagaggg	1020
tctcttttcc cgtgtgcacg tatgtgccag ctacattgtg tgaccaaag actggcatac	1080
tggcaacaga tgtcagtgcg gacgacgcgc aaaaactgct gggtgggctc aaccagcgta	1140
tagtcgtcaa cggtcgcacc cagagaaaca ccaataccat gaaaaattac cttttgccccg	1200
tagtggccca ggcatttgct aggtgggcaa aggaatataa ggaagatcaa gaagatgaaa	1260
ggccactagg actacgagat agacagttag tcatgggggtg ttgttgggct tttagaaggc	1320
acaagataac atctatttat aagcggccgg atacccaaac catcatcaaa gtgaacagcg	1380
atttcactc attcgtgctg cccaggatag gcagtaacac attggagatc gggctgagaa	1440
caagaatcag gaaaatgtta gaggagcaca aggagccgtc acctctcatt accgccgagg	1500
acgtacaaga agctaagtgc gcagccgatg aggctaagga ggtgcgtgaa gccgaggagt	1560
tgcgcgcagc tctaccacct ttggcagctg atgttgagga gcccactctg gaagccgatg	1620
tcgacttgat gttacaagag gctggggccg gctcagtgga gacacctcgt ggcttgataa	1680
aggttaccag ctacgctggc gaggacaaga tcggctctta cgctgtgctt tctccgcagg	1740

ctgtactcaa	gagtgaaaaa	ttatcttgca	tccaccctct	cgctgaacaa	gtcatagtga	1800
taacacactc	tggccgaaaa	gggcgttatg	ccgtggaacc	ataccatggg	aaagtagtgg	1860
tgccagaggg	acatgcaata	cccgtccagg	actttcaagc	tctgagtga	agtgccacca	1920
ttgtgtacaa	cgaacgtgag	ttcgtaaaca	ggtacctgca	ccatattgcc	acacatggag	1980
gagcgctgaa	cactgatgaa	gaatattaca	aaactgtcaa	gcccagcgag	cacgacggcg	2040
aatacctgta	cgacatcgac	aggaaacagt	gcgtcaagaa	agaactagtc	actgggctag	2100
ggctcacagg	cgagctgggtg	gaccttcct	tccatgaatt	cgcttacgag	agtctgagaa	2160
cacgaccagc	cgctccttac	caagtaccaa	ccataggggt	gtatggcgtg	ccaggatcag	2220
gcaagtctgg	catcattaaa	agcgcagtca	caaaaaaga	tctagtgggtg	agcgccaaga	2280
aagaaaactg	tgcagaaatt	ataagggacg	tcaagaaaat	gaaagggtg	gacgtcaatg	2340
ccagaactgt	ggactcagt	ctcttgaatg	gatgcaaaca	ccccgtagag	accctgtata	2400
ttgacgaagc	ttttgcttgt	catgcaggta	ctctcagagc	gctcatagcc	attataagac	2460
ctaaaaaggc	agtgtctctgc	ggggatccca	aacagtgcgg	tttttttaac	atgatgtgcc	2520
tgaaagtgca	ttttaaccac	gagatttgca	cacaagtctt	ccacaaaagc	atctctcgcc	2580
gttgactaa	atctgtgact	tcggtcgtct	caaccttggt	ttacgacaaa	aaaatgagaa	2640
cgacgaatcc	gaaagagact	aagattgtga	ttgacactac	cggcagtacc	aaacctaagc	2700
aggacgatct	cattctcact	tgtttcagag	ggtgggtgaa	gcagttgcaa	atagattaca	2760
aaggcaacga	aataatgacg	gcagctgcct	ctcaagggt	gacccgtaaa	ggtgtgtatg	2820
ccgttcggta	caaggtgaat	gaaaatcctc	tgtacgcacc	cacctcagaa	catgtgaacg	2880
tcctactgac	ccgcacggag	gaccgcatcg	tgtggaaaac	actagccggc	gacccatgga	2940
taaaaacact	gactgccaag	taccctggga	atttcactgc	cacgatagag	gagtggcaag	3000
cagagcatga	tgccatcatg	aggcacatct	tggagagacc	ggaccctacc	gacgtcttcc	3060
agaataaggc	aaacgtgtgt	tgggccaagg	ctttagtgcc	ggtgctgaag	accgctggca	3120
tagacatgac	cactgaacaa	tggaacactg	tggattattt	tgaaacggac	aaagctcact	3180
cagcagagat	agtattgaac	caactatgcg	tgaggttctt	tggactcgat	ctggactccg	3240
gtctattttc	tgcaccact	gttccgttat	ccattaggaa	taatcactgg	gataactccc	3300
cgtcgcctaa	catgtacggg	ctgaataaag	aagtgggtccg	tcagctctct	cgcaggtacc	3360

cacaactgcc tcgggcagtt gccactggaa gagtctatga catgaacact ggtacactgc	3420
gcaattatga tccgcgcata aacctagtag ctgtaaacag aagactgcct catgctttag	3480
tcctccacca taatgaacac ccacagagtg actttttcttc attcgtcagc aaattgaagg	3540
gcagaactgt cctgggtggtc ggggaaaagt tgtccgtccc aggcaaaatg gttgactggt	3600
tgtcagaccg gcctgaggct accttcagag ctctggctgga tttaggcatc ccaggatgatg	3660
tgcccaaata tgacataata tttgttaatg tgaggacccc atataaatac catcactatc	3720
agcagtgtga agaccatgcc attaagctta gcatgttgac caagaaagct tgtctgcatc	3780
tgaatcccgg cggaacctgt gtcagcatag gttatgggta cgctgacagg gccagcgaaa	3840
gcatcattgg tgctatagcg cggcagttca agttttcccg ggtatgcaaa ccgaaatcct	3900
cacttgaaga gacggaagtt ctgtttgtat tcattgggta cgatcgcaag gcccgtacgc	3960
acaatcctta caagctttca tcaaccttga ccaacattta tacagggttcc agactccacg	4020
aagccggatg tgcaccctca tatcatgttg tgcgagggga tattgccacg gccaccgaag	4080
gagtgattat aaatgctgct aacagcaaag gacaacctgg cggaggggtg tgcggagcgc	4140
tgtataagaa attcccggaa agcttcgatt tacagccgat cgaagtagga aaagcgcgac	4200
tggtcaaagg tgcagctaaa catatcattc atgccgtagg accaaacttc aacaaagttt	4260
cggaggttga aggtgacaaa cagttggcag aggcttatga gtccatcgct aagattgtca	4320
acgataacaa ttacaagtca gtagcgattc cactgttggtc caccggcatc ttttccggga	4380
acaaagatcg actaaccctaa tcattgaacc atttgctgac agcttttagac accactgatg	4440
cagatgtagc catatactgc agggacaaga aatgggaaat gactctcaag gaagcagtgg	4500
ctaggagaga agcagtggag gagatatgca tatccgacga ctcttcagtg acagaacctg	4560
atgcagagct ggtgaggggtg catccgaaga gttctttggc tggaagggaag ggctacagca	4620
caagcgtatg caaaactttc tcataatttg aagggaacca gtttcaccag gcggccaagg	4680
atatagcaga aattaatgcc atgtggcccg ttgcaacgga ggccaatgag caggatatgca	4740
tgtatatcct cggagaaaagc atgagcagta ttaggtcgaa atgccccgtc gaagagtcgg	4800
aagcctccac accacctagc acgctgcctt gcttgtgcat ccatgccatg actccagaaa	4860
gagtacagcg cctaaaagcc tcacgtccag aacaaattac tgtgtgctca tcctttccat	4920
tgccgaagta tagaatcact ggtgtgcaga agatccaatg ctcccagcct atattgttct	4980

caccgaaagt gcctgcgtat attcatccaa ggaagtatct cgtggaaaca ccaccggtag	5040
acgagactcc ggagccatcg gcagagaacc aatccacaga ggggacacct gaacaaccac	5100
cacttataac cgaggatgag accaggacta gaacgcctga gccgatcatc atcgaagagg	5160
aagaagagga tagcataagt ttgctgtcag atggcccgac ccaccagggtg ctgcaagtcg	5220
aggcagacat tcacgggccg ccctctgtat ctagctcatc ctggtccatt cctcatgcat	5280
ccgactttga tgtggacagt ttatccatac ttgacacctt ggagggagct agcgtgacca	5340
gcggggcaac gtcagccgag actaactctt acttcgcaaa gagtatggag tttctggcgc	5400
gaccggtgcc tgcgcctcga acagtattca ggaaccctcc acatcccgct ccgcgcacaa	5460
gaacaccgtc acttgacccc agcagggcct gctcgagaac cagcctagtt tccaccccgc	5520
caggcgtgaa taggggtgatc actagagagg agctcgaggc gcttaccctg tcacgcactc	5580
ctagcaggtc ggtctcgaga accagcctgg tctccaacct gccaggcgta aatagggtga	5640
ttacaagaga ggagtttgag gcgttcgtag cacaacaaca atgacggttt gatgcgggtg	5700
catacatctt ttcctccgac accggtcaag ggcatttaca acaaaaatca gtaaggcaaa	5760
cggtgctatc cgaagtgggtg ttggagagga ccgaattgga gatttcgtat gccccgcgcc	5820
tcgaccaaga aaaagaagaa ttactacgca agaaattaca gttaaattccc acacctgcta	5880
acagaagcag ataccagtcc aggaagggtgg agaacatgaa agccataaca gctagacgta	5940
ttctgcaagg cctagggcat tatttgaagg cagaaggaaa agtggagtgc taccgaacct	6000
tgcatcctgt tcctttgtat tcacttagtg tgaaccgtgc cttttcaagc cccaaggctg	6060
cagtggaagc ctgtaacgcc atgttgaaag agaactttcc gactgtggct tcttactgta	6120
ttattccaga gtacgatgcc tatttggaca tggttgacgg agcttcatgc tgcttagaca	6180
ctgccagttt ttgccctgca aagctgcgca gctttccaaa gaaacactcc tatttggaac	6240
ccacaatacg atcggcagtg ccttcagcga tccagaacac gctccagaac gtcctggcag	6300
ctgccacaaa aagaaattgc aatgtcacgc aaatgagaga attgcccgtg ttggattcgg	6360
cggcctttta tgtggaatgc ttcaagaaat atgcgtgtaa taatgaatat tgggaaacgt	6420
ttaaagaaaa ccccatcagg cttactgaag aaaacgtggg aaattacatt accaaattaa	6480
aaggacaaa agctgctgct ctttttgcca agacacataa tttgaatatg ttgcaggaca	6540
taccaatgga caggtttgta atggacttaa agagagacgt gaaagtgact ccaggaacaa	6600

aacatactga	agaacggccc	aaggtacagg	tgatccaggc	tgccgatccg	ctagcaacag	6660
cgtatctgtg	cggaatccac	cgagagctgg	ttaggagatt	aaatgcggtc	ctgcttccga	6720
acattcatac	actgtttgat	atgtcggctg	aagactttga	cgctattata	gccgagcact	6780
tccagcctgg	ggattgtggt	ctggaaactg	acatcgcgtc	gtttgataaa	agtgaggacg	6840
acgccatggc	tctgaccgcg	ttaatgattc	tggaagactt	aggtgtggac	gcagagctgt	6900
tgacgctgat	tgaggcggct	ttcggcgaaa	tttcatcaat	acatttgccc	actaaaacta	6960
aatttaaatt	cggagccatg	atgaaatctg	gaatgttcct	cacactgttt	gtgaacacag	7020
tcattaacat	tgtaatcgca	agcagagtgt	tgagagaacg	gctaaccgga	tcaccatgtg	7080
cagcattcat	tgagatgac	aatatcgtga	aaggagtcaa	atcggacaaa	ttaatggcag	7140
acaggtgcg	cacctggttg	aatatggaag	tcaagattat	agatgctgtg	gtgggcgaga	7200
aagcgcctta	tttctgtgga	gggtttat	tgtgtgactc	cgtgaccggc	acagcgtgcc	7260
gtgtggcaga	ccccctaaaa	aggctgttta	agcttgga	acctctggca	gcagacgatg	7320
aacatgatga	tgacaggaga	agggcattgc	atgaagagtc	aacacgctgg	aaccgagtgg	7380
gtattctttc	agagctgtgc	aaggcagtag	aatcaaggta	tgaaaccgta	ggaacttcca	7440
tcatagtatt	ggccatgact	actctagcta	gcagtgttaa	atcattcagc	tacctgagag	7500
gggcccctat	aactctctac	ggctaacctg	aatggactac	gacatagtct	agtccgcca	7560
gatgttcccg	ttccagccaa	tgtatccgat	gcagccaatg	ccctatcgca	acccgttcgc	7620
ggccccgcgc	aggccctggt	tccccagaac	cgaccctttt	ctggcgatgc	aggtgcagga	7680
attaacccgc	tcgatggcta	acctgacgtt	caagcaacgc	cgggacgcgc	cacctgaggg	7740
gccatccgct	aagaaaccga	agaaggaggc	ctcgcaaaaa	cagaaagggg	gaggccaagg	7800
gaagaagaag	aagaaccaag	ggaagaagaa	ggctaagaca	gggccgccta	atccgaaggc	7860
acagaatgga	aacaagaaga	agaccaacaa	gaaaccaggc	aagagacagc	gcatgggtcat	7920
gaaattggaa	tctgacaaga	cgttcccaat	catgttggaa	gggaagataa	acggctacgc	7980
ttgtgtggtc	ggagggaagt	tattcaggcc	gatgcatgtg	gaaggcaaga	tcgacaacga	8040
cgttctggcc	gcgcttaaga	cgaagaaagc	atccaaatac	gatcttgagt	atgcagatgt	8100
gccacagaac	atgcgggccg	atacattcaa	atacaccat	gagaaacccc	aaggctatta	8160
cagctggcat	catggagcag	tccaatatga	aaatgggcgt	ttcacgggtgc	cgaaaggagt	8220

tggggccaag ggagacagcg gacgacccat tctggataac cagggacggg tggtcgctat	8280
tgtgctggga ggtgtgaatg aaggatctag gacagccctt tcagtcgtca tgtggaacga	8340
gaagggagtt accgtgaagt atactccgga gaactgcgag caatggtcac tagtgaccac	8400
catgtgtctg ctcgccaatg tgacgttccc atgtgctcaa ccaccaattt gctacgacag	8460
aaaaccagca gagactttgg ccatgctcag cgттаacgtt gacaacccgg gctacgatga	8520
gctgctggaa gcagctgtta agtgccccgg aaggaaaagg agatccaccg aggagctgtt	8580
taaggagtat aagctaacgc gcccttacat ggccagatgc atcagatgtg cagttggggag	8640
ctgccatagt ccaatagcaa tcgaggcagt aaagagcgac gggcacgacg gttatgttag	8700
acttcagact tcctcgcagt atggcctgga ttcttcggc aacttaaagg gcaggaccat	8760
gcggtatgac atgcacggga ccattaaaga gataccacta catcaagtgt cactccatac	8820
atctcgcccc tgtcacattg tggatgggca cggttatttc ctgcttgcca ggtgccccggc	8880
aggggactcc atcaccatgg aatttaagaa agattccgtc acacactcct gctcggtgcc	8940
gtatgaagtg aaatttaatc ctgtaggcag agaactctat actcatcccc cagaacacgg	9000
agtagagcaa gcgtgccaaг tctacgcaca tgatgcacag aacagaggag cttatgtcga	9060
gatgcacctc ccgggctcag aagtggacag cagtttggtt tccttgagcg gcagttcagt	9120
caccgtgaca cctcctgttg ggactagcgc cctggtggaa tgcgagtgtg gcggcacaaa	9180
gatctccgag accatcaaca agacaaaaca gttcagccag tgcacaaaga aggagcagtг	9240
cagagcatat cggctgcaga acgataagtг ggtgtataat tctgacaaac tgcccaaagc	9300
agcgggagcc accttaaaag gaaaactgca tgtcccattc ttgctggcag acggcaaатг	9360
caccgtgcct ctagcaccag aacctatgat aacctttggt ttcagatcag tgtcactgaa	9420
actgcaccct aagaatccca catatctaac caccgccea cttgctgatg agcctcacta	9480
cacgcacgag ctcatatctg aaccagctgt taggaatttt accgtcaccg aaaaagggtг	9540
ggagtttgta tggggaaacc acccgccgaa aaggttttgg gcacaggaaa cagcaccggg	9600
aaatccacat gggctaccgc acgaggtgat aactcattat taccacagat accctatgtc	9660
caccatcctg ggtttgtcaa tttgtgccgc cattgcaacc gtttccgttg cagcgtctac	9720
ctggctgttt tgcagatcta gagttgcgtg cctaactcct taccggctaa cacctaacgc	9780
taggatacca ttttgtctgg ctgtgctttg ctgcgccccг actgcccggg ccgagaccac	9840

ctgggagtcc ttggatcacc tatggaacaa taaccaacag atgttctgga ttcaattgct	9900
gatccctctg gccgccttga tcgtagtgac tcgcctgctc aggtgctgt gctgtgtcgt	9960
gcctttttta gtcattggccg gcgccgcagg cgccggcgcc tacgagcacg cgaccacgat	10020
gccgagccaa gcgggaatct cgtataacac tatagtcaac agagcaggct acgcaccact	10080
ccctatcagc ataacaccaa caaagatcaa gctgatacct acagtgaact tggagtacgt	10140
cacctgccac taaaaaacag gaatggattc accagccatc aaatgctgcg gatctcagga	10200
atgcactcca acttacaggc ctgatgaaca gtgcaaagtc ttcacagggg tttacccgtt	10260
catgtggggg ggtgcatatt gcttttgcga cactgagaac acccaagtca gcaaggccta	10320
cgtaatgaaa tctgacgact gccttgcgga tcatgctgaa gcatataaag cgcacacagc	10380
ctcagtgcag gcgttcctca acatcacagt gggagaacac tctattgtga ctaccgtgta	10440
tgtgaatgga gaaactcctg tgaatttcaa tgggggtcaaa ttaactgcag gtccgctttc	10500
cacagcttgg acaccctttg atcgcaaaat cgtgcagtat gccggggaga tctataatta	10560
tgattttcct gagtatgggg caggacaacc aggagcattt ggagatatac aatccagaac	10620
agtctcaagc tcagatctgt atgccaatac caacctagtg ctgcagagac ccaaagcagg	10680
agcgatccac gtgccataca ctcaggcacc ttcgggtttt gagcaatgga agaaagataa	10740
agctccatca ttgaaattta cgcgcccttt cggatgcgaa atatatacaa accccattcg	10800
cgccgaaaac tgtgctgtag ggtcaattcc attagccttt gacattcccg acgccttggt	10860
caccagggtg tcagaaacac cgacactttc agcggccgaa tgcactctta acgagtgcgt	10920
gtattcttcc gactttggtg ggatcgccac ggtcaagtac tcggccagca agtcaggcaa	10980
gtgcgcagtc catgtgcat cagggactgc taccctaaaa gaagcagcag tcgagctaac	11040
cgagcaaggg tcggcgacta tccattttct gaccgcaa atccacccgg agttcaggct	11100
ccaaatatgc acatcatatg ttacgtgcaa aggtgattgt cccccccga aagaccatat	11160
tgtgacacac cctcagtatc acgccccaa atttacagcc gcggtgtcaa aaaccgcgtg	11220
gacgtgggta acatccctgc tgggaggatc agccgtaatt attataattg gcttggtgct	11280
ggctactatt gtggccatgt acgtgctgac caaccagaaa cataattgaa tacagcagca	11340
attggcaagc tgcttacata gaactcgcgg cgattggcat gccgccttaa aatttttatt	11400
ttattttttc ttttcttttc cgaatcggat tttgttttta atatttc	11447

<210> 4
<211> 9577
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
polynucleotide

<400> 4
atgggcggcg catgagagaa gcccagacca attacctacc caaaatggag aaagttcacg 60
ttgacatcga ggaagacagc ccattcctca gagctttgca gcggagcttc ccgcagtttg 120
aggtagaagc caagcaggtc actgataatg accatgctaa tgccagagcg ttttcgcatc 180
tggcttcaaa actgatcgaa acggaggtgg acccatccga cacgatcctt gacattggaa 240
gtgcgcccgc ccgcagaatg tattctaagc acaagtatca ttgtatctgt ccgatgagat 300
gtgcggaaga tccggacaga ttgtataagt atgcaactaa gctgaagaaa aactgtaagg 360
aaataactga taaggaattg gacaagaaaa tgaaggagct cgccgccgtc atgagcgacc 420
ctgacctgga aactgagact atgtgcctcc acgacgacga gtcgtgtcgc tacgaagggc 480
aagtcgctgt ttaccaggat gtatacgcgg ttgacggacc gacaagtctc taccaccaag 540
ccaataaggg agttagagtc gcctactgga taggctttga caccaccctt tttatgttta 600
agaacttggc tggagcatat ccatacact ctaccaactg ggccgacgaa accgtgttaa 660
cggctcgtaa cataggccta tgcagctctg acgttatgga gcggtcacgt agagggatgt 720
ccattcttag aaagaagtat ttgaaaccat ccaacaatgt tctattctct gttggctcga 780
ccatctacca cgagaagagg gacttactga ggagctggca cctgccgtct gtatttcact 840
tacgtggcaa gcaaaattac acatgtcggg gtgagactat agttagttgc gacgggtacg 900
tcgttaaaag aatagctatc agtccaggcc tgtatgggaa gccttcaggc tatgctgcta 960
cgatgcaccg cgagggattc ttgtgctgca aagtgacaga cacattgaac ggggagaggg 1020
tctcttttcc cgtgtgcacg tatgtgccag ctacattgtg tgaccaaag actggcatac 1080
tggcaacaga tgtcagtgcg gacgacgcgc aaaaactgct gggtgggctc aaccagcgta 1140
tagtcgtcaa cggctgcacc cagagaaaca ccaataccat gaaaaattac cttttgccccg 1200
tagtggccca ggcatttgct aggtgggcaa aggaatataa ggaagatcaa gaagatgaaa 1260
ggccactagg actacgagat agacagttag tcatgggggtg ttgttgggct tttagaaggc 1320

acaagataac atctatztat aagcgcccgg ataccctaac catcatcaaa gtgaacagcg	1380
atttccactc attcgtgctg cccaggatag gcagtaacac attggagatc gggctgagaa	1440
caagaatcag gaaaatgtta gaggagcaca aggagccgtc acctctcatt accgccgagg	1500
acgtacaaga agctaagtgc gcagccgatg aggctaagga ggtgcgtgaa gccgaggagt	1560
tgcgcgcagc tctaccacct ttggcagctg atgttgagga gccactctg gaagccgatg	1620
tcgacttgat gttacaagag gctggggccg gctcagtga gacacctcgt ggcttgataa	1680
aggttaccag ctacgctggc gaggacaaga tcggctctta cgctgtgctt tctccgcagg	1740
ctgtactcaa gagtgaacaa ttatcttgca tccacctctt cgctgaacaa gtcatagtga	1800
taacacactc tggccgaaaa gggcgttatg ccgtggaacc ataccatggg aaagtagtgg	1860
tgccagaggg acatgcaata cccgtccagg actttcaagc tctgagtga agtgccacca	1920
ttgtgtacaa cgaacgtgag ttcgtaacaa ggtacctgca ccatattgcc acacatggag	1980
gagcgtgaa cactgatgaa gaattattaca aaactgtcaa gccagcgag cacgacggcg	2040
aatacctgta cgacatcgac aggaacagc gcgtcaagaa agaactagtc actgggctag	2100
ggctcacagg cgagctgggt gatcctccct tccatgaatt cgcctacgag agtctgagaa	2160
cacgaccagc cgctccttac caagtaccaa ccataggggt gtatggcgtg ccaggatcag	2220
gcaagtctgg catcattaaa agcgcagtca ccaaaaaaga tctagtgggt agcgccaaga	2280
aagaaaactg tgcagaaatt ataaggagc tcaagaaaat gaaagggtg gacgtcaatg	2340
ccagaactgt ggactcagt ctcttgaatg gatgcaaaca ccccgtagag accctgtata	2400
ttgacgaagc ttttgcttgt catgcaggta ctctcagagc gctcatagcc attataagac	2460
ctaaaaaggc agtgctctgc ggggatccca aacagtgcgg tttttttaac atgatgtgcc	2520
tgaaagtgca ttttaaccac gagatttgca cacaagtctt ccacaaaagc atctctcgcc	2580
gttgactaa atctgtgact tcggtcgtct caaccttggt ttacgacaaa aaaatgagaa	2640
cgacgaatcc gaaagagact aagattgtga ttgacactac cggcagtacc aaacctaagc	2700
aggacgatct cattctcact tgtttcagag ggtgggtgaa gcagttgcaa atagattaca	2760
aaggcaacga aataatgacg gcagctgcct ctcaagggtg gaccgtaaa ggtgtgtatg	2820
ccgttcggta caaggtgaat gaaaatcctc tgtacgcacc cacctcagaa catgtgaacg	2880
tcctactgac ccgcacggag gaccgcatcg tgtggaaaac actagccggc gaccatgga	2940

taaaaacact gactgccaag taccctggga atttctactgc cacgatagag gagtggcaag	3000
cagagcatga tgccatcatg aggcacatct tggagagacc ggaccttacc gacgtcttcc	3060
agaataaggc aaacgtgtgt tgggccaagg ctttagtgcc ggtgctgaag accgctggca	3120
tagacatgac cactgaacaa tggaacactg tggattatct tgaaacggac aaagctcact	3180
cagcagagat agtattgaac caactatgcg tgaggttctt tggactcgat ctggactccg	3240
gtctattttc tgcaccctact gttccgttat ccattaggaa taatcactgg gataactccc	3300
cgtcgcctaa catgtacggg ctgaataaag aagtgggtccg tcagctctct cgaggtacc	3360
cacaactgcc tcgggcagtt gccactggaa gagtctatga catgaacact ggtacactgc	3420
gcaattatga tccgcgcata aacctagtac ctgtaaacag aagactgcct catgctttag	3480
tcctccacca taatgaacac ccacagagtg acttttcttc attcgtcagc aaattgaagg	3540
gcagaactgt cctggtgggtc ggggaaaagt tgtccgtccc aggcaaatg gttgactggt	3600
tgtcagaccg gcctgaggct accttcagag ctcggctgga tttaggcatc ccaggtgatg	3660
tgcccaaata tgacataata tttgttaatg tgaggacccc atataaatac catcactatc	3720
agcagtgtga agaccatgcc attaagctta gcatgttgac caagaaagct tgtctgcac	3780
tgaatcccgg cggaacctgt gtcagcatag gttatggtta cgctgacagg gccagcga	3840
gcatcattgg tgctatagcg cggcagttca agttttcccg ggtatgcaaa ccgaaatcct	3900
cacttgaaga gacggaagtt ctgtttgtat tcattgggta cgatcgcaag gcccgtacgc	3960
acaatcctta caagctttca tcaaccttga ccaacattta tacagggtcc agactccacg	4020
aagccggatg tgcaccctca tatcatgttg tgcgagggga tattgccacg gccaccgaag	4080
gagtgattat aaatgctgct aacagcaaag gacaacctgg cggaggggtg tgcggagcgc	4140
tgtataagaa attcccggaa agcttcgatt tacagccgat cgaagtagga aaagcgcgac	4200
tggtcaaagg tgcagctaaa catatcattc atgccgtagg accaaacttc aacaaagttt	4260
cggaggttga aggtgacaaa cagttggcag aggcttatga gtccatcgct aagattgtca	4320
acgataacaa ttacaagtca gtagcgattc cactgttgct caccggcatc ttttccggga	4380
acaaagatcg actaaccctaa tcattgaacc atttgctgac agcttttagac accactgatg	4440
cagatgtagc catatactgc agggacaaga aatgggaaat gactctcaag gaagcagtgg	4500
ctaggagaga agcagtggag gagatatgca tatccgacga ctcttcagtg acagaacctg	4560

atgcagagct ggtgagggtg catccgaaga gttctttggc tggaaggaag ggctacagca	4620
caagcgatgg caaaactttc tcatatttgg aagggaacaa gtttcaccag gcggccaagg	4680
atatagcaga aattaatgcc atgtggcccg ttgcaacgga ggccaatgag caggtatgca	4740
tgtatatcct cggagaaagc atgagcagta ttaggtcgaa atgccccgtc gaagagtcgg	4800
aagcctccac accacctagc acgctgcctt gcttgtgcat ccatgccatg actccagaaa	4860
gagtacagcg cctaaaagcc tcacgtccag aacaaattac tgtgtgctca tcctttccat	4920
tgccgaagta tagaatcact ggtgtgcaga agatccaatg ctcccagcct atattgttct	4980
caccgaaagt gcctgcgtat attcatccaa ggaagtatct cgtggaaaca ccaccggtag	5040
acgagactcc ggagccatcg gcagagaacc aatccacaga ggggacacct gaacaaccac	5100
cacttataac cgaggatgag accaggacta gaacgcctga gccgatcatc atcgaagagg	5160
aagaagagga tagcataagt ttgctgtcag atggcccgac ccaccagggtg ctgcaagtcg	5220
aggcagacat tcacgggccg ccctctgtat ctagctcatc ctgggtccatt cctcatgcat	5280
ccgactttga tgtggacagt ttatccatac ttgacacctt ggaggggagct agcgtgacca	5340
gcggggcaac gtcagccgag actaactctt acttcgcaaa gagtatggag tttctggcgc	5400
gaccggtgcc tgcgcctcga acagtattca ggaaccctcc acatcccgtc ccgcgcacaa	5460
gaacaccgtc acttgcaccc agcagggcct gctcgagaac cagcctagtt tccaccccgc	5520
caggcgtgaa taggggtgac actagagagg agctcgaggc gcttaccctg tcacgcactc	5580
ctagcaggtc ggtctcgaga accagcctgg tctccaacct gccaggcgta aatagggtga	5640
ttacaagaga ggagtttgag gcgttcgtag cacaacaaca atgacggttt gatgcgggtg	5700
catacatctt ttcctccgac accggtcaag ggcatttaca aaaaaatca gtaaggcaaa	5760
cggtgctatc cgaagtgggtg ttggagagga ccgaattgga gatttcgtat gccccgcgcc	5820
tcgaccaaga aaaagaagaa ttactacgca agaaattaca gttaaattccc acacctgcta	5880
acagaagcag ataccagtcc aggaagggtg agaacatgaa agccataaca gctagacgta	5940
ttctgcaagg cctagggcat tatttgaagg cagaaggaaa agtggagtgc taccgaacct	6000
tgcatcctgt tcctttgtat tcacttagtg tgaaccgtgc cttttcaagc cccaaggctg	6060
cagtggaagc ctgtaacgcc atgttgaaag agaactttcc gactgtggct tcttactgta	6120
ttattccaga gtacgatgcc tatttggaca tggttgacgg agcttcatgc tgcttagaca	6180

ctgccagttt ttgccctgca aagctgcgca gctttccaaa gaaacactcc tatttggaac	6240
ccacaatacg atcggcagtg ccttcagcga tccagaacac gctccagaac gtcctggcag	6300
ctgccacaaa aagaaattgc aatgtcacgc aaatgagaga attgcccgtg ttggattcgg	6360
cggcctttta tgtggaatgc ttcaagaaat atgcgtgtaa taatgaatat tgggaaacgt	6420
ttaaagaaaa ccccatcagg cttactgaag aaaacgtggg aaattacatt accaaattaa	6480
aaggacaaa agctgctgct ctttttgcg agacacataa tttgaatatg ttgcaggaca	6540
taccaatgga caggtttgta atggacttaa agagagacgt gaaagtgact ccaggaacaa	6600
aacatactga agaacggccc aaggtacagg tgatccaggc tgccgatccg ctagcaacag	6660
cgtatctgtg cggaatccac cgagagctgg ttaggagatt aaatgcggtc ctgcttccga	6720
acattcatac actgtttgat atgtcggctg aagactttga cgctattata gccgagcact	6780
tccagcctgg ggattgtgtt ctggaaactg acatcgcgtc gtttgataaa agtgaggacg	6840
acgccatggc tctgaccgcg ttaatgattc tggaagactt aggtgtggac gcagagctgt	6900
tgacgctgat tgaggcggct ttcggcgaaa tttcatcaat acatttgccc actaaaacta	6960
aatttaaat cggagccatg atgaaatctg gaatgttcct cacactgttt gtgaacacag	7020
tcattaacat tgtaatcgca agcagagtgt tgagagaacg gctaaccgga tcaccatgtg	7080
cagcattcat tggagatgac aatatcgtga aaggagtcaa atcggacaaa ttaatggcag	7140
acaggtgctg cacctggttg aatatggaag tcaagattat agatgctgtg gtgggcgaga	7200
aagcgcctta tttctgtgga gggtttattt tgtgtgactc cgtgaccggc acagcgtgcc	7260
gtgtggcaga cccctaaaa aggctgttta agcttgcaa acctctggca gcagacgatg	7320
aacatgatga tgacaggaga agggcattgc atgaagagtc aacacgctgg aaccgagtgg	7380
gtattctttc agagctgtgc aaggcagtag aatcaaggta tgaaaccgta ggaacttcca	7440
tcatagtatt ggccatgact actctagcta gcagtgttaa atcattcagc tacctgagag	7500
gggcccctat aactctctac ggctaacctg aatggactac gactctagaa tagtctttaa	7560
ttaagccacc atggcaggca tgtttcaggc gctgagcgaa ggctgcaccc cgtatgatat	7620
taaccagatg ctgaacgtgc tgggcgatca tcaggtctca ggcccttgagc agcttgagag	7680
tataatcaac tttgaaaaac tgactgaatg gaccagttct aatgttatgc ctatcctgtc	7740
tcctctgaca aagggcattc tgggcttcgt gtttacctg accgtgcctt ctgagagagg	7800

acttagctgc attagcgaag cggatgcgac cccccggaa agcgcgaaacc tgggcgaaga	7860
aattctgagc cagctgtatc tttggccaag ggtgacctac cattccccta gttatgctta	7920
ccaccaattt gaaagacgag ccaaataataa aagacacttc cccggccttg gccagagcct	7980
gctgtttggc taccctgtgt acgtgttcgg cgattgcgtg cagggcgatt gggatgcgat	8040
tcgctttcgc tattgcgcgc cgccgggcta tgcgtgctg cgctgcaacg ataccaacta	8100
tagcgctctg ctggctgtgg gggccctaga aggaccagc aatcaggact ggcttggtgt	8160
ccaagacaa cttgtaactc ggatgcaggc tattcagaat gccggcctgt gtaccctggt	8220
ggccatgctg gaagagacaa tcttctggct gcaagcgttt ctgatggcgc tgaccgatag	8280
cggcccga aa accaacatta ttgtggatag ccagtatgtg atgggcatta gcaaaccgag	8340
ctttcaggaa tttgtggatt gggaaaacgt gagcccggaa ctgaacagca ccgatcagcc	8400
gttttgcaa gccggaatcc tggccagaaa tctggtgcct atggtggcca cagtgcaggg	8460
ccagaacctg aagtaccagg gtcagtcact agtcatctct gcttctatca ttgtcttcaa	8520
cctgctggaa ctggaagggtg attatcgaga tgatggcaac gtgtgggtgc ataccccgct	8580
gagcccgcgc accctgaacg cgtgggtgaa agcgggtggaa gaaaaaaaaag gtattccagt	8640
tcacctagag ctggccagta tgaccaacat ggagctcatg agcagtattg tgcacagca	8700
ggtcagaaca tacggccccg tgttcatgtg tctcggcgga ctgcttaciaa tgggtggctgg	8760
tgctgtgtgg ctgacagtgc gagtgctcga gctgttccgg gccgcgcagc tggccaacga	8820
cgtggtcctc cagatcatgg agctttgtgg tgcagcgttt cgccagggtgt gccataccac	8880
cgtgccgtgg ccgaacgcga gcctgacccc gaaatggaac aacgaaacca cccagcccca	8940
gatcgccaac tgcagcgtgt atgacttttt tgtgtggctc cattattatt ctgttcgaga	9000
cacactttgg ccaagggtga cctaccatat gaacaaatat gcgtatcata tgctggaaag	9060
acgagccaaa tataaaagag gaccaggacc tggcgctaaa tttgtggccg cctggacact	9120
gaaagccgct gctggtcctg gacctggcca gtacatcaag gccaacagca agttcatcgg	9180
catcaccgaa ctcggacccg gaccaggctg atgattcgaa cggccgtatc acgcccac	9240
atttacagcc gcggtgtcaa aaaccgcgtg gacgtgggta acatccctgc tgggaggatc	9300
agccgtaatt attataattg gcttggtgct ggctactatt gtggccatgt acgtgctgac	9360
caaccagaaa cataattgaa tacagcagca attggcaagc tgcttacata gaactcgcgg	9420

cgattggcat gccgccttaa aatTTTTtatt ttatTTTTTtc ttttcttttc cgaatcggat	9480
tttgttttta atatttcaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa	9540
aaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaa	9577

<210> 5
 <211> 11447
 <212> DNA
 <213> Venezuelan equine encephalitis virus

<400> 5	
atgggcggcg catgagagaa gcccagacca attacctacc caaaatggag aaagttcacg	60
ttgacatcga ggaagacagc ccatttctca gagctttgca gcggagcttc ccgcagtttg	120
aggtagaagc caagcaggtc actgataatg accatgctaa tgccagagcg ttttcgcatc	180
tggcttcaaa actgatcgaa acggagggtgg acccatccga cacgatcctt gacattggaa	240
gtgcgcccgc ccgcagaatg tattctaagc acaagtatca ttgtatctgt ccgatgagat	300
gtgcggaaga tccggacaga ttgtataagt atgcaactaa gctgaagaaa aactgtaagg	360
aaataactga taaggaattg gacaagaaaa tgaaggagct cgccgccgtc atgagcgacc	420
ctgacctgga aactgagact atgtgcctcc acgacgacga gtcgtgtcgc tacgaagggc	480
aagtcgctgt ttaccaggat gtatacgcggt ttgacggacc gacaagtctc tatcaccaag	540
ccaataaggg agttagagtc gcctactgga taggctttga caccaccctt tttatgttta	600
agaacttggc tggagcatat ccatcactat ctaccaactg ggccgacgaa accgtgttaa	660
cggctcgtaa cataggccta tgcagctctg acgttatgga gcggtcacgt agagggatgt	720
ccattcttag aaagaagtat ttgaaaccat ccaacaatgt tctattctct gttggctcga	780
ccatctacca cgagaagagg gacttactga ggagctggca cctgccgtct gtatttcact	840
tacgtggcaa gcaaaattac acatgtcggt gtgagactat agttagttgc gacgggtacg	900
tcgttaaaaag aatagctatc agtccaggcc tgtatgggaa gccttcaggc tatgctgcta	960
cgatgcaccg cgagggattc ttgtgctgca aagtgcagca cacattgaac ggggagaggg	1020
tctcttttcc cgtgtgcacg tatgtgccag ctacattgtg tgaccaaatg actggcatac	1080
tggcaacaga tgtcagtgcg gacgacgcgc aaaaactgct gggtgggctc aaccagcgta	1140
tagtcgtcaa cggtcgcacc cagagaaaca ccaataccat gaaaaattac cttttgcccg	1200

tagtggccca	ggcatttgct	aggtgggcaa	aggaatataa	ggaagatcaa	gaagatgaaa	1260
ggccactagg	actacgagat	agacagttag	tcatgggggtg	ttgttgggct	tttagaaggc	1320
acaagataac	atctatttat	aagcgcccgg	atacccaaac	catcatcaaa	gtgaacagcg	1380
atttccactc	attcgtgctg	cccaggatag	gcagtaacac	attggagatc	gggctgagaa	1440
caagaatcag	gaaaatgtta	gaggagcaca	aggagccgtc	acctctcatt	accgccgagg	1500
acgtacaaga	agctaagtgc	gcagccgatg	aggctaagga	ggtgcgtgaa	gccgaggagt	1560
tgcgcgcagc	tctaccacct	ttggcagctg	atgttgagga	gcccactctg	gaagccgatg	1620
tcgacttgat	gttacaagag	gctggggccg	gctcagtgga	gacacctcgt	ggcttgataa	1680
aggttaccag	ctacgctggc	gaggacaaga	tcggctctta	cgctgtgctt	tctccgcagg	1740
ctgtactcaa	gagtgaaaaa	ttatcttgca	tccacctctt	cgctgaacaa	gtcatagtga	1800
taacacactc	tggccgaaaa	gggcgttatg	ccgtggaacc	ataccatggg	aaagtagtgg	1860
tgccagaggg	acatgcaata	cccgtccagg	actttcaagc	tctgagtgaa	agtgccacca	1920
tttgtgtaca	cgaacgtgag	ttcgtaaaca	ggtacctgca	ccatattgcc	acacatggag	1980
gagcgtgaa	cactgatgaa	gaatattaca	aaactgtcaa	gcccagcgag	cacgacggcg	2040
aatacctgta	cgacatcgac	aggaaacagt	gcgtcaagaa	agaactagtc	actgggctag	2100
ggctcacagg	cgagctgggt	gaccttcctt	tccatgaatt	cgcctacgag	agtctgagaa	2160
cacgaccagc	cgctccttac	caagtaccaa	ccataggggt	gtatggcgtg	ccaggatcag	2220
gcaagtctgg	catcattaaa	agcgcagtca	ccaaaaaaga	tctagtgggt	agcgccaaga	2280
aagaaaactg	tgcagaaatt	ataagggacg	tcaagaaaat	gaaagggctg	gacgtcaatg	2340
ccagaactgt	ggactcagtg	ctcttgaatg	gatgcaaaca	ccccgtagag	accctgtata	2400
ttgacgaagc	ttttgcttgt	catgcaggta	ctctcagagc	gctcatagcc	attataagac	2460
ctaaaaaggc	agtgctctgc	ggggatccca	aacagtgcgg	tttttttaac	atgatgtgcc	2520
tgaaagtgca	ttttaaccac	gagatttgca	cacaagtctt	ccacaaaagc	atctctcgcc	2580
gttgactact	atctgtgact	tcggtcgtct	caaccttggt	ttacgacaaa	aaaatgagaa	2640
cgacgaatcc	gaaagagact	aagattgtga	ttgacactac	cggcagtacc	aaacctaagc	2700
aggacgatct	cattctcact	tgtttcagag	ggtgggtgaa	gcagttgcaa	atagattaca	2760
aaggcaacga	aataatgacg	gcagctgcct	ctcaagggtc	gacccgtaaa	ggtgtgtatg	2820

ccgttcggta caaggtgaat gaaaatcctc tgtacgcacc cacctcagaa catgtgaacg	2880
tcctactgac ccgcacggag gaccgcatcg tgtggaaaac actagccggc gacccatgga	2940
taaaaaact gactgccaaag taccctggga atttcactgc cacgatagag gagtggcaag	3000
cagagcatga tgccatcatg aggcacatct tggagagacc ggaccctacc gacgtcttcc	3060
agaataaggc aaacgtgtgt tgggccaaagg ctttagtgcc ggtgctgaag accgctggca	3120
tagacatgac cactgaacaa tggaacactg tggattatit tgaaacggac aaagctcact	3180
cagcagagat agtattgaac caactatgcg tgaggttctt tggactcgat ctggactccg	3240
gtctatititc tgcaccact gttccgttat ccattaggaa taatcactgg gataactccc	3300
cgctgcctaa catgtacggg ctgaataaag aagtgggtccg tcagctctct cgcaggtacc	3360
cacaactgcc tcgggcagtt gccactggaa gagtctatga catgaacact ggtacactgc	3420
gcaattatga tccgcgcata aacctagtac ctgtaaacag aagactgcct catgctttag	3480
tcctccacca taatgaacac ccacagagtg acttttcttc attcgtcagc aaattgaagg	3540
gcagaactgt cctggtggtc ggggaaaagt tgtccgtccc aggcaaatg gttgactggt	3600
tgtcagaccg gcctgaggct accttcagag ctcggttgga tttaggcatc ccaggtgatg	3660
tgcccaaata tgacataata tttgttaatg tgaggacccc atataaatac catcactatc	3720
agcagtgtga agaccatgcc attaaagctta gcatgttgac caagaaagct tgtctgcac	3780
tgaatcccgg cggaacctgt gtcagcatag gttatggtta cgctgacagg gccagcgaaa	3840
gcatcattgg tgctatagcg cggcagttca agttttcccg ggtagtcaaa ccgaaatcct	3900
cacttgaaga gacggaagtt ctgtttgtat tcattgggta cgatcgcaag gcccgtagc	3960
acaatcctta caagctttca tcaaccttga ccaacattta tacaggttcc agactccacg	4020
aagccggatg tgcaccctca tatcatgttg tgcgagggga tattgccacg gccaccgaag	4080
gagtgattat aaatgctgct aacagcaaag gacaacctgg cggaggggtg tgcggagcgc	4140
tgtataagaa attcccggaa agcttcgatt tacagccgat cgaagtagga aaagcgcgac	4200
tggtcaaagg tgcagctaaa catatcattc atgccgtagg accaaacttc aacaaagttt	4260
cggaggttga aggtgacaaa cagttggcag aggcttatga gtccatcgct aagattgtca	4320
acgataacaa ttacaagtca gtagcgattc cactgttgct caccggcatc ttttccggga	4380
acaaagatcg actaaccaa tcattgaacc atttgctgac agcttttagac accactgatg	4440

cagatgtagc catatactgc agggacaaga aatgggaaat gactctcaag gaagcagtgg	4500
ctaggagaga agcagtggag gagatatgca tatccgacga ctcttcagtg acagaacctg	4560
atgcagagct ggtgaggggtg catccgaaga gttctttggc tggaaggaag ggctacagca	4620
caagcgatgg caaaactttc tcatatttgg aagggaacca gtttcaccag gcggccaagg	4680
atatagcaga aattaatgcc atgtggcccg ttgcaacgga ggccaatgag caggtatgca	4740
tgtatatacct cggagaaagc atgagcagta ttaggtcgaa atgccccgtc gaagagtcgg	4800
aagcctccac accacctagc acgctgcctt gcttgtgcat ccatgccatg actccagaaa	4860
gagtacagcg cctaaaagcc tcacgtccag aacaaattac tgtgtgctca tcctttccat	4920
tgccgaagta tagaatcact ggtgtgcaga agatccaatg ctcccagcct atattgttct	4980
caccgaaagt gcctgcgtat attcatccaa ggaagtatct cgtggaaaca ccaccggtag	5040
acgagactcc ggagccatcg gcagagaacc aatccacaga ggggacacct gaacaaccac	5100
cacttataac cgaggatgag accaggacta gaacgcctga gccgatcatc atcgaagagg	5160
aagaagagga tagcataagt ttgtgtcag atggcccgac ccaccagggtg ctgcaagtcg	5220
aggcagacat tcacgggccg ccctctgtat ctagctcatc ctggtccatt cctcatgcat	5280
ccgactttga tgtggacagt ttatccatac ttgacaccct ggaggggagct agcgtgacca	5340
gcggggcaac gtcagccgag actaactctt acttcgcaaa gagtatggag tttctggcgc	5400
gaccggtgcc tgcgcctcga acagtattca ggaaccctcc acatcccgtc ccgcgcaaca	5460
gaacaccgtc acttgcaccc agcagggcct gctcgagaac cagcctagtt tccaccccg	5520
caggcgtgaa tagggatgatc actagagagg agctcgaggc gcttaccggtc tcacgcactc	5580
ctagcaggtc ggtctcgaga accagcctgg tctccaacct gccaggcgta aatagggtga	5640
ttacaagaga ggagtttgag gcgttcgtag cacaacaaca atgacgggtt gatgcgggtg	5700
catacatctt ttcctccgac accgggtcaag ggcatattaca acaaaaatca gtaaggcaaa	5760
cggtgctatc cgaagtgggtg ttggagagga ccgaattgga gatttcgtat gccccgcgcc	5820
tcgaccaaga aaaagaagaa ttactacgca agaaattaca gttaaattccc acacctgcta	5880
acagaagcag ataccagtcc aggaagggtg agaacatgaa agccataaca gctagacgta	5940
ttctgcaagg cctagggcat tatttgaagg cagaaggaaa agtggagtgc taccgaacct	6000
tgcatacctgt tcctttgtat tcatactagt tgaaccgtgc cttttcaagc cccaaggctg	6060

cagtggaagc ctgtaacgcc atgttgaaag agaactttcc gactgtggct tcttactgta	6120
ttattccaga gtacgatgcc tatttggaca tggttgacgg agcttcatgc tgcttagaca	6180
ctgccagttt ttgccctgca aagctgcgca gctttccaaa gaaacactcc tatttggaac	6240
ccacaatacg atcggcagtg ctttcagcga tccagaacac gctccagaac gtcctggcag	6300
ctgccacaaa aagaaattgc aatgtcacgc aaatgagaga attgcccgtg ttggattcgg	6360
cggcctttta tgtggaatgc ttcaagaaat atgcgtgtaa taatgaatat tgggaaacgt	6420
ttaaagaaaa ccccatcagg cttactgaag aaaacgtggg aaattacatt accaaattaa	6480
aaggacaaa agctgctgct ctttttgcga agacacataa tttgaatatg ttgcaggaca	6540
taccaatgga caggtttgta atggacttaa agagagacgt gaaagtgact ccaggaacaa	6600
aacatactga agaacggccc aaggtacagg tgatccaggc tgccgatccg ctagcaacag	6660
cgtatctgtg cggaatccac cgagagctgg ttaggagatt aaatgcggtc ctgcttccga	6720
acattcatac actgtttgat atgtcggctg aagactttga cgctattata gccgagcact	6780
tccagcctgg ggattgtgtt ctggaaactg acatcgcgtc gtttgataaa agtgaggacg	6840
acgccatggc tctgaccgcg ttaatgattc tggaagactt aggtgtggac gcagagctgt	6900
tgacgctgat tgaggcggct ttcggcgaaa tttcatcaat acatttgccc actaaaacta	6960
aatttaaatt cggagccatg atgaaatctg gaatgttcct cacactgttt gtgaacacag	7020
tcattaacat tgtaatcgca agcagagtgt tgagagaacg gctaaccgga tcaccatgtg	7080
cagcattcat tggagatgac aatatcgtga aaggagtcaa atcggacaaa ttaatggcag	7140
acaggtgcg cacttggttg aatatggaag tcaagattat agatgctgtg gtgggcgaga	7200
aagcgcctta tttctgtgga gggtttattt tgtgtgactc cgtgaccggc acagcgtgcc	7260
gtgtggcaga cccctaaaa aggctgttta agcttggtcaa acctctggca gcagacgatg	7320
aacatgatga tgacaggaga agggcattgc atgaagagtc aacacgctgg aaccgagtgg	7380
gtattctttc agagctgtgc aaggcagtag aatcaaggta tgaaaccgta ggaacttcca	7440
tcatagttat ggccatgact actctagcta gcagtgttaa atcattcagc tacctgagag	7500
gggcccctat aactctctac ggctaacctg aatggactac gacatagtct agtccgccaa	7560
gatgttcccg ttccagccaa tgtatccgat gcagccaatg ccctatcgca acccgttcgc	7620
ggccccgcgc aggcctgggt tccccagaac cgaccctttt ctggcgatgc aggtgcagga	7680

attaacccgc tcgatggcta acctgacgtt caagcaacgc cgggacgcgc cacctgaggg	7740
gccatccgct aagaaaccga agaaggaggc ctcgcaaaaa cagaaagggg gaggccaagg	7800
gaagaagaag aagaaccaag ggaagaagaa ggctaagaca gggccgccta atccgaaggc	7860
acagaatgga aacaagaaga agaccaacaa gaaaccaggc aagagacagc gcatggtcat	7920
gaaattggaa tctgacaaga cgttcccaat catgttggaa gggaagataa acggctacgc	7980
ttgtgtggtc ggagggaagt tattcaggcc gatgcatgtg gaaggcaaga tcgacaacga	8040
cgttctggcc gcgcttaaga cgaagaaagc atccaaatac gatcttgagt atgcagatgt	8100
gccacagaac atgcgggccg atacattcaa atacacccat gagaaacccc aaggctatta	8160
cagctggcat catggagcag tccaatatga aaatgggcgt ttcacgggtc cgaaaggagt	8220
tggggccaag ggagacagcg gacgacccat tctggataac cagggacggg tggctgctat	8280
tgtgctggga ggtgtgaatg aaggatctag gacagccctt tcagtcgtca tgtggaacga	8340
gaagggagtt accgtgaagt atactccgga gaactgcgag caatggtcac tagtgaccac	8400
catgtgtctg ctgc caatg tgacgttccc atgtgctcaa ccaccaattt gctacgacag	8460
aaaaccagca gagactttgg ccatgctcag cgттаacgtt gacaacccgg gctacgatga	8520
gctgctggaa gcagctgtta agtgccccgg aaggaaaagg agatccaccg aggagctgtt	8580
taaggagtat aagctaacgc gcccttacat ggccagatgc atcagatgtg cagttgggag	8640
ctgccatagt ccaatagcaa tcgaggcagt aaagagcgac gggcacgacg gttatgttag	8700
acttcagact tcctcgcagt atggcctgga ttctcgcggc aacttaaagg gcaggaccat	8760
gcggtatgac atgcacggga ccattaaaga gataccacta catcaagtgt cactccatac	8820
atctcgcccc tgtcacattg tggatgggca cggttatttc ctgcttgcca ggtgccccgc	8880
aggggactcc atcaccatgg aatttaagaa agattccgtc acacactcct gctcggtgcc	8940
gtatgaagtg aaatttaatc ctgtaggcag agaactctat actcatcccc cagaacacgg	9000
agtagagcaa gcgtgccaaag tctacgcaca tgatgcacag aacagaggag cttatgtcga	9060
gatgcacctc ccgggctcag aagtggacag cagtttggtt tccttgagcg gcagttcagt	9120
caccgtgaca cctcctgttg ggactagcgc cctggtggaa tgcgagtgtg gcggcacaaa	9180
gatctccgag accatcaaca agacaaaaca gttcagccag tgcacaaaga aggagcagtg	9240
cagagcatat cggctgcaga acgataagtg ggtgtataat tctgacaaac tgcccaaagc	9300

agcgggagcc	accttaaaag	gaaaactgca	tgtcccattc	ttgctggcag	acggcaaattg	9360
caccgtgcct	ctagcaccag	aacctatgat	aaccttttgg	ttcagatcag	tgtcactgaa	9420
actgcaccct	aagaatccca	catatctaac	caccgcgcaa	cttgctgatg	agcctcacta	9480
cacgcacgag	ctcatatctg	aaccagctgt	taggaatttt	accgtcaccg	aaaaagggtg	9540
ggagtttgta	tggggaaacc	acccgccgaa	aaggttttgg	gcacaggaaa	cagcaccg	9600
aaatccacat	gggctaccgc	acgaggtgat	aactcattat	taccacagat	accctatgtc	9660
caccatcctg	ggtttgtcaa	tttgtgccgc	cattgcaacc	gtttccgttg	cagcgtctac	9720
ctggctgttt	tgcagatcta	gagttgcgtg	cctaactcct	taccggctaa	cacctaacgc	9780
taggatacca	ttttgtctgg	ctgtgctttg	ctgcgccgc	actgcccggg	ccgagaccac	9840
ctgggagtcc	ttggatcacc	tatggaacaa	taaccaacag	atgttctgga	ttcaattgct	9900
gatccctctg	gccgccttga	tcgtagtga	tcgcctgctc	aggtgcgtgt	gctgtgtcgt	9960
gcctttttta	gtcatggccg	gcgccgcagg	cgccggcgcc	tacgagcacg	cgaccacgat	10020
gccgagccaa	gcgggaatct	cgtataacac	tatagtcaac	agagcaggct	acgcaccact	10080
ccctatcagc	ataacaccaa	caaagatcaa	gctgatacct	acagtgaact	tggagtacgt	10140
cacctgccac	tacaaaacag	gaatggattc	accagccatc	aaatgctgcg	gatctcagga	10200
atgcactcca	acttacaggc	ctgatgaaca	gtgcaaagtc	ttcacagggg	tttaccggtt	10260
catgtggggg	ggtgcatatt	gcttttgcga	cactgagaac	acccaagtca	gcaaggccta	10320
cgtaatgaaa	tctgacgact	gccttgcgga	tcatgttgaa	gcatataaag	cgcacacagc	10380
ctcagtgcag	gcgttcctca	acatcacagt	gggagaacac	tctattgtga	ctaccgtgta	10440
tgtgaatgga	gaaactcctg	tgaatttcaa	tgggggtcaaa	ttactgcag	gtccgctttc	10500
cacagcttgg	acaccctttg	atcgcaaaat	cgtgcagtat	gccggggaga	tctataatta	10560
tgattttcct	gagtatgggg	caggacaacc	aggagcattt	ggagatatac	aatccagaac	10620
agtctcaagc	tcagatctgt	atgccaatac	caacctagt	ctgcagagac	ccaaagcagg	10680
agcgatccac	gtgccataca	ctcaggcacc	ttcgggtttt	gagcaatgga	agaaagataa	10740
agctccatca	ttgaaattta	ccgccccttt	cggatgcgaa	atatatacaa	accccatctg	10800
cgccgaaaac	tgtgctgtag	ggtcaattcc	attagccttt	gacattccc	acgccttggt	10860
caccagggtg	tcagaaacac	cgacactttc	agcggccgaa	tgcactctta	acgagtgcgt	10920

gtattcttcc gactttggtg ggatcgccac ggtcaagtac tcggccagca agtcaggcaa	10980
gtgcgcagtc catgtgccat cagggactgc taccctaaaa gaagcagcag tcgagctaac	11040
cgagcaaggg tcggcgacta tccatttctc gaccgcaa atccacccgg agttcaggct	11100
ccaaatatgc acatcatatg ttacgtgcaa aggtgattgt cccccccga aagaccatat	11160
tgtgacacac cctcagtatc acgccccaaac atttacagcc gcggtgtcaa aaaccgcgtg	11220
gacgtggtta acatccctgc tgggaggatc agccgtaatt attataattg gcttgggtgct	11280
ggctactatt gtggccatgt acgtgctgac caaccagaaa cataattgaa tacagcagca	11340
attggcaagc tgcttacata gaactcgcgg cgattggcat gccgccttaa aatTTTTatt	11400
ttattttttc ttttcttttc cgaatcggat tttgttttta atatttc	11447

<210> 6

<211> 7894

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 6

atgggcggcg catgagagaa gcccagacca attacctacc caaaatggag aaagttcacg	60
ttgacatcga ggaagacagc ccatttctca gagctttgca gcggagcttc ccgcagtttg	120
aggtagaagc caagcaggtc actgataatg accatgctaa tgccagagcg ttttcgcatc	180
tggcttcaaa actgatcgaa acggagggtg acccatccga cacgatcctt gacattggaa	240
gtgcgcccgc ccgcagaatg tattctaagc acaagtatca ttgtatctgt ccgatgagat	300
gtgcggaaga tccggacaga ttgtataagt atgcaactaa gctgaagaaa aactgtaagg	360
aaataactga taaggaattg gacaagaaaa tgaaggagct cgccgccgtc atgagcgacc	420
ctgacctgga aactgagact atgtgcctcc acgacgacga gtcgtgtcgc tacgaagggc	480
aagtcgctgt ttaccaggat gtatacgcgg ttgacggacc gacaagtctc tatcaccaag	540
ccaataaggg agttagagtc gcctactgga taggctttga caccaccctt tttatgttta	600
agaacttggc tggagcatat ccatcact ctaccaactg ggccgacgaa accgtgttaa	660
cggctcgtaa cataggccta tgcagctctg acgttatgga gcggtcacgt agagggatgt	720
ccattcttag aaagaagtat ttgaaaccat ccaacaatgt tctattctct gttggctcga	780

ccatctacca	cgagaagagg	gacttactga	ggagctggca	cctgccgtct	gtatttcact	840
tacgtggcaa	gcaaaattac	acatgtcggg	gtgagactat	agttagttgc	gacgggtacg	900
tcgttaaaag	aatagctatc	agtccaggcc	tgtatgggaa	gccttcaggc	tatgctgcta	960
cgatgcaccg	cgagggattc	ttgtgctgca	aagtgacaga	cacattgaac	ggggagaggg	1020
tctcttttcc	cgtgtgcacg	tatgtgccag	ctacattgtg	tgaccaaata	actggcatac	1080
tggcaacaga	tgtcagtgcg	gacgacgcgc	aaaaactgct	ggttgggctc	aaccagcgta	1140
tagtcgtcaa	cggtcgcacc	cagagaaaca	ccaataccat	gaaaaattac	cttttgcccc	1200
tagtggccca	ggcatttgct	aggtggggcaa	aggaatataa	ggaagatcaa	gaagatgaaa	1260
ggccactagg	actacgagat	agacagttag	tcatgggggtg	ttgttgggct	tttagaaggc	1320
acaagataac	atctatttat	aagcgcccgg	atacccaaac	catcatcaaa	gtgaacagcg	1380
atttcactc	attcgtgctg	cccaggatag	gcagtaacac	attggagatc	gggctgagaa	1440
caagaatcag	gaaaatgtta	gaggagcaca	aggagccgtc	acctctcatt	accgccgagg	1500
acgtacaaga	agctaagtgc	gcagccgatg	aggctaagga	ggtgcgtgaa	gccgaggagt	1560
tgcgcgcagc	tctaccacct	ttggcagctg	atgttgagga	gcccactctg	gaagccgatg	1620
tcgacttgat	gttacaagag	gctggggccg	gctcagtgga	gacacctcgt	ggcttgataa	1680
aggttaccag	ctacgctggc	gaggacaaga	tcggctctta	cgctgtgctt	tctccgcagg	1740
ctgtactcaa	gagtgaaaaa	ttatcttgca	tccaccctct	cgctgaacaa	gtcatagtga	1800
taacacactc	tggccgaaaa	gggcgttatg	ccgtggaacc	ataccatggg	aaagtagtgg	1860
tgccagaggg	acatgcaata	cccgtccagg	actttcaagc	tctgagtga	agtgccacca	1920
tttgtgtaca	cgaacgtgag	ttcgtaaaca	ggtacctgca	ccatattgcc	acacatggag	1980
gagcgctgaa	cactgatgaa	gaatattaca	aaactgtcaa	gcccagcgag	cacgacggcg	2040
aatacctgta	cgacatcgac	aggaaacagt	gcgtcaagaa	agaactagtc	actgggctag	2100
ggctcacagg	cgagctgggtg	gacacctcct	tccatgaatt	cgcttacgag	agtctgagaa	2160
cacgaccagc	cgctccttac	caagtaccaa	ccataggggt	gtatggcgtg	ccaggatcag	2220
gcaagtctgg	catcattaaa	agcgcagtca	ccaaaaaaga	tctagtgggtg	agcgccaaga	2280
aagaaaactg	tgcagaaatt	ataagggacg	tcaagaaaat	gaaagggtg	gacgtcaatg	2340
ccagaactgt	ggactcagtg	ctcttgaatg	gatgcaaaca	ccccgtagag	accctgtata	2400

ttgacgaagc	ttttgcttgt	catgcaggta	ctctcagagc	gctcatagcc	attataagac	2460
ctaaaaaggc	agtgctctgc	ggggatccca	aacagtgcgg	tttttttaac	atgatgtgcc	2520
tgaaagtgca	ttttaaccac	gagatttgca	cacaagtctt	ccacaaaagc	atctctcgcc	2580
gttgcactaa	atctgtgact	tcggtcgtct	caaccttggt	ttacgacaaa	aaaatgagaa	2640
cgacgaatcc	gaaagagact	aagattgtga	ttgacactac	cggcagtacc	aaacctaagc	2700
aggacgatct	cattctcact	tgtttcagag	ggtgggtgaa	gcagttgcaa	atagattaca	2760
aaggcaacga	aataatgacg	gcagctgcct	ctcaagggtc	gacccgtaaa	ggtgtgtatg	2820
ccgttcggta	caaggtgaat	gaaaatcctc	tgtacgcacc	cacctcagaa	catgtgaacg	2880
tcctactgac	ccgcacggag	gaccgcatcg	tgtggaaaac	actagccggc	gacccatgga	2940
taaaaacact	gactgccaag	taccctggga	atttcactgc	cacgatagag	gagtggcaag	3000
cagagcatga	tgccatcatg	aggcacatct	tggagagacc	ggaccctacc	gacgtcttcc	3060
agaataaggc	aaacgttgtg	tgggccaaag	ctttagtgcc	ggtgctgaag	accgctggca	3120
tagacatgac	cactgaacaa	tggaacactg	tggattatit	tgaaacggac	aaagctcact	3180
cagcagagat	agtattgaac	caactatgcg	tgaggttctt	tggactcgat	ctggactccg	3240
gtctatitit	tgcacccact	gttccgttat	ccattaggaa	taatcactgg	gataactccc	3300
cgtcgcctaa	catgtacggg	ctgaataaag	aagtgggtcc	tcagctctct	cgcaggtacc	3360
cacaactgcc	tcgggcagtt	gccactggaa	gagtctatga	catgaacact	ggtacactgc	3420
gcaattatga	tccgcgcata	aacctagtac	ctgtaaacag	aagactgcct	catgctttag	3480
tcctccacca	taatgaacac	ccacagagtg	acttttcttc	attcgtcagc	aaattgaagg	3540
gcagaactgt	cctggtggtc	ggggaaaagt	tgtccgtccc	aggcaaaatg	gttgactggt	3600
tgtcagaccg	gcctgaggct	accttcagag	ctcggctgga	tttaggcata	ccaggtgatg	3660
tgcccaaata	tgacataata	tttgttaatg	tgaggacccc	atataaatac	catcactatc	3720
agcagtgtga	agaccatgcc	attaagctta	gcatgttgac	caagaaagct	tgtctgcata	3780
tgaatcccgg	cggaacctgt	gtcagcatag	gttatggtta	cgctgacagg	gccagcgaaa	3840
gcatcattgg	tgctatagcg	cggcagttca	agttttcccg	ggtatgcaaa	ccgaaatcct	3900
cacttgaaga	gacggaagtt	ctgtttgtat	tcattgggta	cgatcgcaag	gcccgtacgc	3960
acaatcctta	caagctttca	tcaaccttga	ccaacattta	tacaggttcc	agactccacg	4020

aagccggatg tgcaccctca tatcatgtgg tgcgagggga tattgccacg gccaccgaag	4080
gagtgattat aaatgctgct aacagcaaag gacaacctgg cggaggggtg tgcggagcgc	4140
tgtataagaa attcccggaa agcttcgatt tacagccgat cgaagtagga aaagcgcgac	4200
tgggtcaaagg tgcagctaaa catatcattc atgccgtagg accaaacttc aacaaagttt	4260
cggaggttga aggtgacaaa cagttggcag aggcttatga gtccatcgct aagattgtca	4320
acgataacaa ttacaagtca gtagcgattc cactgttgtc caccggcatc ttttccggga	4380
acaaagatcg actaaccaa tcattgaacc atttgctgac agctttagac accactgatg	4440
cagatgtagc catatactgc agggacaaga aatgggaaat gactctcaag gaagcagtgg	4500
ctaggagaga agcagtggag gagatatgca tatccgacga ctcttcagtg acagaacctg	4560
atgcagagct ggtgaggggtg catccgaaga gttctttggc tggaaggaag ggctacagca	4620
caagcgatgg caaaactttc tcatatttgg aagggaacaa gtttcaccag gcggccaagg	4680
atatagcaga aattaatgcc atgtggcccg ttgcaacgga ggccaatgag caggtatgca	4740
tgtatatacct cggagaaagc atgagcagta ttaggtcgaa atgccccgtc gaagagtcgg	4800
aagcctccac accacctagc acgctgcctt gcttgtgcat ccatgccatg actccagaaa	4860
gagtacagcg cctaaaagcc tcacgtccag aacaaattac tgtgtgctca tcctttccat	4920
tgccgaagta tagaatcact ggtgtgcaga agatccaatg ctcccagcct atattgttct	4980
caccgaaagt gcctgcgtat attcatcaa ggaagtatct cgtggaaaca ccaccggtag	5040
acgagactcc ggagccatcg gcagagaacc aatccacaga ggggacacct gaacaaccac	5100
cacttataac cgaggatgag accaggacta gaacgcctga gccgatcatc atcgaagagg	5160
aagaagagga tagcataagt ttgtgtcag atggcccgcac ccaccagggtg ctgcaagtcg	5220
aggcagacat tcacgggccg ccctctgtat ctagctcatc ctggtccatt cctcatgcat	5280
ccgactttga tgtggacagt ttatccatac ttgacaccct ggaggggagct agcgtgacca	5340
gcggggcaac gtcagccgag actaactctt acttcgcaaa gagtatggag tttctggcgc	5400
gaccggtgcc tgcgcctcga acagtattca ggaaccctcc acatcccgtc ccgcgacaaa	5460
gaacaccgtc acttgcaccc agcagggcct gctcgagaac cagcctagtt tccaccccg	5520
caggcgtgaa tagggtgatc actagagagg agctcgaggc gcttaccgag tcacgcactc	5580
ctagcaggtc ggtctcgaga accagcctgg tctccaaccc gccaggcgta aatagggtga	5640

ttacaagaga	ggagtttgag	gcgttcgtag	cacaacaaca	atgacggttt	gatgcgggtg	5700
catacatctt	ttcctccgac	accggtcaag	ggcattttaca	acaaaaatca	gtaaggcaaa	5760
cgggtgctatc	cgaagtgggtg	ttggagagga	ccgaattgga	gatttcgtat	gccccgcgcc	5820
tcgaccaaga	aaaagaagaa	ttactacgca	agaaattaca	gttaaattccc	acacctgcta	5880
acagaagcag	ataccagtcc	aggaagggtg	agaacatgaa	agccataaca	gctagacgta	5940
ttctgcaagg	cctagggcat	tatttgaagg	cagaaggaaa	agtggagtgc	taccgaaccc	6000
tgcatcctgt	tcctttgtat	tcatctagt	tgaaccgtgc	cttttcaagc	cccaaggctg	6060
cagtggaagc	ctgtaacgcc	atgttgaaag	agaactttcc	gactgtggct	tcttactgta	6120
ttattccaga	gtacgatgcc	tatttggaca	tggttgacgg	agcttcatgc	tgcttagaca	6180
ctgccagttt	ttgccctgca	aagctgcgca	gctttccaaa	gaaacactcc	tatttggaac	6240
ccacaatacg	atcggcagtg	ccttcagcga	tccagaacac	gctccagaac	gtcctggcag	6300
ctgccacaaa	aagaaattgc	aatgtcacgc	aaatgagaga	attgcccgtg	ttggattcgg	6360
cggcctttta	tgtggaatgc	ttcaagaaat	atgcgtgtaa	taatgaatat	tgggaaacgt	6420
ttaaagaaaa	ccccatcagg	cttactgaag	aaaacgtgg	aaattacatt	accaaattaa	6480
aaggacaaa	agctgctgct	ctttttgcga	agacacataa	tttgaatatg	ttgcaggaca	6540
taccaatgga	caggtttgta	atggacttaa	agagagacgt	gaaagtgact	ccaggaacaa	6600
aacatactga	agaacggccc	aaggtacagg	tgatccaggc	tgccgatccg	ctagcaacag	6660
cgtatctgtg	cggaatccac	cgagagctgg	ttaggagatt	aaatgcggtc	ctgcttccga	6720
acattcatac	actgtttgat	atgtcggctg	aagactttga	cgctattata	gccgagcact	6780
tccagcctgg	ggattgtgtt	ctggaaactg	acatcgcgtc	gtttgataaa	agtgaggacg	6840
acgccatggc	tctgaccgcg	ttaatgattc	tggaagactt	aggtgtggac	gcagagctgt	6900
tgacgctgat	tgaggcggct	ttcggcgaaa	tttcatcaat	acatttgccc	actaaaacta	6960
aatttaaat	cggagccatg	atgaaatctg	gaatgttcct	cacactgttt	gtgaacacag	7020
tcattaacat	tgtaatcgca	agcagagtgt	tgagagaacg	gctaaccgga	tcacatgtg	7080
cagcattcat	tggagatgac	aatatcgtga	aaggagtcaa	atcggacaaa	ttaatggcag	7140
acagggtgcg	cacctggttg	aatatggaag	tcaagattat	agatgctgtg	gtgggcgaga	7200
aagcgcctta	tttctgtgga	gggtttat	tgtgtgactc	cgtgaccggc	acagcgtgcc	7260

gtgtggcaga cccctaataa aggctgttta agcttggcaa acctctggca gcagacgatg	7320
aacatgatga tgacaggaga agggcattgc atgaagagtc aacacgctgg aaccgagtgg	7380
gtattctttc agagctgtgc aaggcagtag aatcaaggta tgaaaccgta ggaacttcca	7440
tcatagttat ggccatgact actctagcta gcagtgttaa atcattcagc tacctgagag	7500
gggcccctat aactctctac ggctaacctg aatggactac gactatcacg cccaaacatt	7560
tacagccgcg gtgtcaaaaa ccgcgtggac gtgggttaaca tccctgctgg gaggatcagc	7620
cgtaattatt ataattggct tgggtgctggc tactattgtg gccatgtacg tgctgaccaa	7680
ccagaaacat aattgaatac agcagcaatt ggcaagctgc ttacatagaa ctgcggcgga	7740
ttggcatgcc gccttaaaat ttttatttta ttttttcttt tcttttccga atcggatttt	7800
gtttttaata tttcaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa	7860
aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaa	7894

<210> 7

<211> 7893

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic
polynucleotide

<400> 7

ataggcggcg catgagagaa gcccagacca attacctacc caaaatggag aaagttcacg	60
ttgacatcga ggaagacagc ccattcctca gagctttgca gcggagcttc ccgagtttg	120
aggtagaagc caagcaggtc actgataatg accatgctaa tgccagagcg ttttcgcatc	180
tggcttcaaa actgatcgaa acggaggtgg acccatccga cacgatcctt gacattggaa	240
gtgcgcccgc ccgcagaatg tattctaagc acaagtatca ttgtatctgt ccgatgagat	300
gtgcggaaga tccggacaga ttgtataagt atgcaactaa gctgaagaaa aactgtaagg	360
aaataactga taaggaattg gacaagaaaa tgaaggagct cgccgccgtc atgagcgacc	420
ctgacctgga aactgagact atgtgcctcc acgacgacga gtcgtgtcgc tacgaagggc	480
aagtcgctgt ttaccaggat gtatacgcgg ttgacggacc gacaagtctc tatcaccaag	540
ccaataaggg agttagagtc gcctactgga taggctttga caccaccctt tttatgttta	600

agaacttggc tggagcatat ccatcatact ctaccaactg ggccgacgaa accgtgttaa	660
cggctcgtaa cataggccta tgcagctctg acgttatgga gcggtcacgt agagggatgt	720
ccattcttag aaagaagtat ttgaaaccat ccaacaatgt tctattctct gttggctcga	780
ccatctacca cgagaagagg gacttactga ggagctggca cctgccgtct gtatttcact	840
tacgtggcaa gcaaaattac acatgtcggg gtgagactat agttagttgc gacgggtacg	900
tcgttaaaag aatagctatc agtccaggcc tgtatgggaa gccttcaggc tatgctgcta	960
cgatgcaccg cgagggattc ttgtgctgca aagtgcaga cacattgaac ggggagaggg	1020
tctcttttcc cgtgtgcacg tatgtgccag ctacattgtg tgaccaaagc actggcatac	1080
tggcaacaga tgtcagtgcg gacgacgcgc aaaaactgct gggtgggctc aaccagcgta	1140
tagtcgtcaa cggctgcacc cagagaaaca ccaataccat gaaaaattac cttttgccc	1200
tagtggccca ggcatttgct aggtgggcaa aggaatataa ggaagatcaa gaagatgaaa	1260
ggccactagg actacgagat agacagttag tcatgggggtg ttgttgggct tttagaaggc	1320
acaagataac atctatttat aagcgcggc atacccaaac catcatcaaa gtgaacagcg	1380
atttccactc attcgtgctg cccaggatag gcagtaacac attggagatc gggctgagaa	1440
caagaatcag gaaaatgtta gaggagcaca aggagccgtc acctctcatt accgccgagg	1500
acgtacaaga agctaagtgc gcagccgatg aggctaagga ggtgcgtgaa gccgaggagt	1560
tgcgcgcagc tctaccacct ttggcagctg atgttgagga gcccactctg gaagccgatg	1620
tcgacttgat gttacaagag gctggggccg gctcagtgga gacacctcgt ggcttgataa	1680
aggttaccag ctacgatggc gaggacaaga tcggctctta cgctgtgctt tctccgcagg	1740
ctgtactcaa gagtgaaaaa ttatcttgca tccaccctct cgctgaacaa gtcatagtga	1800
taacacactc tggccgaaaa gggcgttatg ccgtggaacc ataccatggg aaagtagtgg	1860
tgccagaggg acatgcaata cccgtccagg actttcaagc tctgagtga agtgccacca	1920
ttgtgtacaa cgaacgtgag ttcgtaaaca ggtacctgca ccatattgcc acacatggag	1980
gagcgtgaa cactgatgaa gaatattaca aaactgtcaa gcccagcgag cacgacggcg	2040
aatacctgta cgacatcgac aggaaacagt gcgtcaagaa agaactagtc actgggctag	2100
ggctcacagg cgagctgggt gatcctccct tccatgaatt cgcctacgag agtctgagaa	2160
cacgaccagc cgctccttac caagtaccaa ccataggggt gtatggcgtg ccaggatcag	2220

gcaagtctgg catcattaa agcgcagtca caaaaaaga tctagtgggtg agcgccaaga	2280
aagaaaactg tgcagaaatt ataagggacg tcaagaaaat gaaagggctg gacgtcaatg	2340
ccagaactgt ggactcagtg ctcttgaatg gatgcaaaca ccccgtagag accctgtata	2400
ttgacgaagc ttttgcttgt catgcaggta ctctcagagc gctcatagcc attataagac	2460
ctaaaaaggc agtgctctgc ggggatccca aacagtgcgg tttttttaac atgatgtgcc	2520
tgaaagtgca ttttaaccac gagatttgca cacaagtctt ccacaaaagc atctctcgcc	2580
gttgactaa atctgtgact tcggtcgtct caaccttggt ttacgacaaa aaaatgagaa	2640
cgacgaatcc gaaagagact aagattgtga ttgacactac cggcagtacc aaacctaacg	2700
aggacgatct cattctcact tgtttcagag ggtgggtgaa gcagttgcaa atagattaca	2760
aaggcaacga aataatgacg gcagctgcct ctcaagggtg gacccgtaaa ggtgtgtatg	2820
ccgttcggta caaggtgaat gaaaatcctc tgtacgcacc cacctcagaa catgtgaacg	2880
tcctactgac ccgcacggag gaccgcatcg tgtggaaaac actagccggc gacccatgga	2940
taaaaacact gactgccaag taccctggga atttactgc cacgatagag gagggtgcaag	3000
cagagcatga tgccatcatg aggcacatct tggagagacc ggaccctacc gacgtcttcc	3060
agaataaggc aaacgtgtgt tgggccaagg ctttagtgcc ggtgctgaag accgctggca	3120
tagacatgac cactgaacaa tggaacactg tggattatct tgaaacggac aaagctcact	3180
cagcagagat agtattgaac caactatgag tgagggtctt tggactcgat ctggactccg	3240
gtctatcttc tgcaccact gttccgttat ccattaggaa taatcactgg gataactccc	3300
cgtcgcctaa catgtacggg ctgaataaag aagtgggtccg tcagctctct cgacaggtacc	3360
cacaactgcc tcgggcagtt gccactggaa gaggctatga catgaacact ggtacactgc	3420
gcaattatga tccgcgcata aacctagtac ctgtaaacag aagactgcct catgctttag	3480
tcctccacca taatgaacac ccacagagtg acttttcttc attcgtcagc aaattgaagg	3540
gcagaactgt cctgggtggc ggggaaaagt tgtccgtccc aggcaaaatg gttgactggt	3600
tgtcagaccg gcctgaggct accttcagag ctgggtgga tttaggcatc ccagggtgatg	3660
tgcccaaata tgacataata tttgttaatg tgaggacccc atataaatac catcactatc	3720
agcagtgtga agaccatgcc attaatgcta gcatgttgac caagaaagct tgtctgcac	3780
tgaatcccg cggaacctgt gtcagcatag gttatgggta cgctgacagg gccagcgaaa	3840

gcatcattgg tgctatagcg cggcagttca agttttcccg ggtatgcaaa ccgaaatcct	3900
cacttgaaga gacggaagtt ctgtttgtat tcattgggta cgatcgcaag gcccgtacgc	3960
acaatcctta caagctttca tcaaccttga ccaacattta tacagggtcc agactccacg	4020
aagccggatg tgcaccctca tatcatgttg tgcgagggga tattgccacg gccaccgaag	4080
gagtgattat aaatgctgct aacagcaaag gacaacctgg cggaggggtg tgcggagcgc	4140
tgtataagaa attccccgaa agcttcgatt tacagccgat cgaagtagga aaagcgcgac	4200
tggtcaaagg tgcagctaaa catatcattc atgccgtagg accaaacttc aacaaagttt	4260
cggaggttga aggtgacaaa cagttggcag aggcttatga gtccatcgct aagattgtca	4320
acgataacaa ttacaagtca gtagcgattc cactgttgtc caccggcatc ttttccggga	4380
acaaagatcg actaaccctca tcattgaacc atttgctgac agcttttagac accactgatg	4440
cagatgtagc catatactgc agggacaaga aatgggaaat gactctcaag gaagcagtgg	4500
ctaggagaga agcagtggag gagatatgca tatccgacga ctcttcagtg acagaacctg	4560
atgcagagct ggtgaggggtg catccgaaga gttctttggc tggaagggaag ggctacagca	4620
caagcgatgg caaaactttc tcatatttgg aaggggacaa gtttcaccag gcggccaagg	4680
atatagcaga aattaatgcc atgtggcccg ttgcaacgga ggccaatgag caggatatgca	4740
tgtatatcct cggagaaagc atgagcagta ttaggtcgaa atgccccgtc gaagagtcgg	4800
aagcctccac accacctagc acgctgcctt gcttgtgcat ccatgccatg actccagaaa	4860
gagtacagcg cctaaaagcc tcacgtccag aacaaattac tgtgtgctca tcctttccat	4920
tgccgaagta tagaatcact ggtgtgcaga agatccaatg ctcccagcct atattgttct	4980
caccgaaagt gcctgcgtat attcatccaa ggaagtatct cgtggaaaca ccaccggtag	5040
acgagactcc ggagccatcg gcagagaacc aatccacaga ggggacacct gaacaaccac	5100
cacttataac cgaggatgag accaggacta gaacgcctga gccgatcatc atcgaagagg	5160
aagaagagga tagcataagt ttgctgtcag atggcccgac ccaccagggtg ctgcaagtcg	5220
aggcagacat tcacggggccg ccctctgtat ctagctcatc ctggtccatt cctcatgcat	5280
ccgactttga tgtggacagt ttatccatac ttgacacctt ggaggggagct agcgtgacca	5340
gcggggcaac gtcagccgag actaactctt acttcgcaaa gagtatggag tttctggcgc	5400
gaccggtgcc tgcgcctcga acagtattca ggaaccctcc acatcccgtc ccgcgcacaa	5460

gaacaccgtc acttgacccc agcagggcct gctcgagaac cagcctagtt tccaccccgc	5520
caggcgtgaa taggggtgatc actagagagg agctcgaggc gcttaccgccg tcacgcactc	5580
ctagcaggtc ggtctcgaga accagcctgg tctccaaccc gccaggcgta aataggggtga	5640
ttacaagaga ggagtttgag gcgttcgtag cacaacaaca atgacgggtt gatgcgggtg	5700
catacatctt ttcctccgac accgggtcaag ggcatttaca acaaaaatca gtaaggcaaa	5760
cgggtgctatc cgaagtgggtg ttggagagga ccgaattgga gatttcgtat gccccgcgcc	5820
tcgaccaaga aaaagaagaa ttactacgca agaaattaca gttaaattccc acacctgcta	5880
acagaagcag ataccagtcc aggaaggtgg agaacatgaa agccataaca gctagacgta	5940
ttctgcaagg cctagggcat tatttgaagg cagaaggaaa agtggagtgc taccgaaccc	6000
tgcattcctgt tcctttgtat tcattctagt tgaaccgtgc cttttcaagc cccaaggctg	6060
cagtggaagc ctgtaacgcc atgttgaaag agaactttcc gactgtggct tcttactgta	6120
ttattccaga gtacgatgcc tatttggaca tggttgacgg agcttcatgc tgcttagaca	6180
ctgccagttt ttgccctgca aagctgcgca gctttccaaa gaaacactcc tatttggaac	6240
ccacaatacg atcggcagtg ccttcagcga tccagaacac gctccagaac gtcctggcag	6300
ctgccacaaa aagaaattgc aatgtcacgc aaatgagaga attgcccgtg ttggattcgg	6360
cggcctttta tgtggaatgc ttcaagaaat atgcgtgtaa taatgaatat tgggaaacgt	6420
ttaaagaaaa ccccatcagg cttactgaag aaaacgtggg aaattacatt accaaattaa	6480
aaggacaaa agctgctgct ctttttgcca agacacataa tttgaatatg ttgcaggaca	6540
taccaatgga caggtttgta atggacttaa agagagacgt gaaagtgact ccaggaacaa	6600
aacatactga agaacggccc aaggtacagg tgatccaggc tgccgatccg ctagcaacag	6660
cgtatctgtg cggaatccac cgagagctgg ttaggagatt aaatgcggtc ctgcttccga	6720
acattcatac actgtttgat atgtcggctg aagactttga cgctattata gccgagcact	6780
tccagcctgg ggattgtgtt ctggaaactg acatcgcgtc gtttgataaa agtgaggacg	6840
acgccatggc tctgaccgag ttaatgattc tggaagactt aggtgtggac gcagagctgt	6900
tgacgctgat tgaggcggct ttcggcgaaa tttcatcaat acatttgccc actaaaacta	6960
aatttaaatt cggagccatg atgaaatctg gaatgttcct cacactgttt gtgaacacag	7020
tcattaacat tgtaatcgca agcagagtgt tgagagaacg gctaaccgga tcaccatgtg	7080

cagcattcat tggagatgac aatatcgtga aaggagtcaa atcggacaaa ttaatggcag	7140
acaggtgcg cacttggttg aatatggaag tcaagattat agatgctgtg gtgggcgaga	7200
aagcgcctta tttctgtgga gggtttattt tgtgtgactc cgtgaccggc acagcgtgcc	7260
gtgtggcaga cccctaaaa aggctgttta agcttggcaa acctctggca gcagacgatg	7320
aacatgatga tgacaggaga agggcattgc atgaagagtc aacacgctgg aaccgagtgg	7380
gtattctttc agagctgtgc aaggcagtag aatcaaggta tgaaaccgta ggaacttcca	7440
tcatagttat ggccatgact actctagcta gcagtgttaa atcattcagc tacctgagag	7500
ggggcccctat aactctctac ggctaacctg aatggactac gactatcacg cccaaacatt	7560
tacagccg cg gtgtcaaaaa ccgcgtggac gtgggttaaca tccctgctgg gaggatcagc	7620
cgtaattatt ataattggct tgggtgctggc tactattgtg gccatgtacg tgctgaccaa	7680
ccagaaacat aattgaatac agcagcaatt ggcaagctgc ttacatagaa ctgcggcgga	7740
ttggcatgcc gccttaaaat ttttatttta tttttctttt cttttccgaa tcggattttg	7800
tttttaatat ttcaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa	7860
aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaa	7893

<210> 8

<211> 7927

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 8

taatacgact cactatagga tgggcggcgc atgagagaag cccagaccaa ttacctacc	60
aaaatggaga aagttcacgt tgacatcgag gaagacagcc cattcctcag agctttgcag	120
cggagcttcc cgcagtttga ggtagaagcc aagcagggtca ctgataatga ccatgcta	180
gccagagcgt tttcgcactt ggcttcaaaa ctgatcgaaa cggagggtgga cccatccgac	240
acgatccttg acattggaag tgcgcccgcc cgcagaatgt attctaagca caagtatcat	300
tgtatctgtc cgatgagatg tgcggaagat ccggacagat tgtataagta tgcaactaag	360
ctgaagaaaa actgtaagga aataactgat aaggaattgg acaagaaaat gaaggagctc	420
gccgccgtca tgagcgaccc tgacctggaa actgagacta tgtgcctcca cgacgacgag	480

tcgtgtcgct	acgaagggca	agtcgctgtt	taccaggatg	tatacgcggt	tgacggaccg	540
acaagtctct	atcaccaagc	caataaggga	gtagagtcg	cctactggat	aggctttgac	600
accacccctt	ttatgtttaa	gaacttggct	ggagcatatc	catcatactc	taccaactgg	660
gccgacgaaa	ccgtgttaac	ggctcgtaac	ataggcctat	gcagctctga	cgttatggag	720
cggtcacgta	gagggatgtc	cattcttaga	aagaagtatt	tgaaaccatc	caacaatgtt	780
ctatttctctg	ttggctcgac	catctaccac	gagaagaggg	acttactgag	gagctggcac	840
ctgccgtctg	tatttcactt	acgtggcaag	caaaattaca	catgtcggtg	tgagactata	900
gtagttgctg	acgggtacgt	cgttaaaaga	atagctatca	gtccaggcct	gtatgggaag	960
ccttcaggct	atgctgctac	gatgcaccgc	gagggattct	tgtgctgcaa	agtgacagac	1020
acattgaacg	gggagagggt	ctcttttccc	gtgtgcacgt	atgtgccagc	tacattgtgt	1080
gaccaaata	ctggcatact	ggcaacagat	gtcagtgcgg	acgacgcgca	aaaactgctg	1140
gttgggctca	accagcgtat	agtcgtcaac	ggtcgcaccc	agagaaacac	caataccatg	1200
aaaaattacc	ttttgcccgt	agtggcccag	gcatttgcta	ggtgggcaaa	ggaatataag	1260
gaagatcaag	aagatgaaag	gccactagga	ctacgagata	gacagttagt	catgggggtgt	1320
tgttgggctt	ttagaaggca	caagataaca	tctatttata	agcgcccgga	tacccaaacc	1380
atcatcaaag	tgaacagcga	tttccactca	ttcgtgctgc	ccaggatagg	cagtaacaca	1440
ttggagatcg	ggctgagaac	aagaatcagg	aaaatgttag	aggagcacia	ggagccgtca	1500
cctctcatta	ccgccgagga	cgtacaagaa	gctaagtgcg	cagccgatga	ggctaaggag	1560
gtgcgtgaag	ccgaggagtt	gcgcgcagct	ctaccacctt	tggcagctga	tgttgaggag	1620
cccactctgg	aagccgatgt	cgacttgatg	ttacaagagg	ctggggccgg	ctcagtggag	1680
acacctctgtg	gcttgataaa	ggttaccagc	tacgctggcg	aggacaagat	cggctcttac	1740
gctgtgcttt	ctccgcaggc	tgtactcaag	agtgaaaaat	tatcttgcat	ccaccctctc	1800
gctgaacaag	tcatagtgat	aacacactct	ggccgaaaag	ggcgttatgc	cgtggaacca	1860
taccatggta	aagtagtggt	gccagaggga	catgcaatac	ccgtccagga	ctttcaagct	1920
ctgagtgaag	gtgccaccat	tgtgtacaac	gaacgtgagt	tcgtaaacag	gtacctgcac	1980
catattgcca	cacatggagg	agcgtgaac	actgatgaag	aatattacaa	aactgtcaag	2040
cccagcgagc	acgacggcga	atacctgtac	gacatcgaca	ggaaacagtg	cgtcaagaaa	2100

gaactagtca ctgggctagg gctcacaggc gagctggtgg atcctccctt ccatgaattc	2160
gcctacgaga gtctgagaac acgaccagcc gtccttacc aagtagcaac cataggggtg	2220
tatggcgtgc caggatcagg caagtctggc atcattaaaa gcgcagtcac caaaaaagat	2280
ctagtgggtga gcgccaagaa agaaaactgt gcagaaatta taaggacgt caagaaaatg	2340
aaagggtg acgtcaatgc cagaactgtg gactcagtc tcttgaatgg atgcaaacac	2400
cccgtagaga ccctgtatat tgacgaagct tttgcttgtc atgcaggtac tctcagagcg	2460
ctcatagcca ttataagacc taaaaaggca gtgctctgcg gggatcccaa acagtgcggt	2520
ttttttaaca tgatgtgcct gaaagtgcac ttttaaccac agatttgcac acaagtcttc	2580
cacaaaagca tctctcgccg ttgcactaaa tctgtgactt cggtcgtctc aaccttgttt	2640
tacgacaaaa aaatgagaac gacgaatccg aaagagacta agattgtgat tgacactacc	2700
ggcagtagca aacctaagca ggacgatctc attctcactt gtttcagagg gtgggtgaag	2760
cagttgcaaa tagattacaa aggcaacgaa ataatgacgg cagctgcctc tcaagggtg	2820
acccgtaaag gtgtgtatgc cggtcgttac aaggtgaatg aaaatcctct gtacgcaccc	2880
acctcagaac atgtgaacgt cctactgacc cgcacggagg accgcatcgt gtggaaaaca	2940
ctagccggcg acccatggat aaaaacactg actgccaagt accctgggaa tttcactgcc	3000
acgatagagg agtggcaagc agagcatgat gccatcatga ggcacatctt ggagagaccg	3060
gaccctaccg acgtcttcca gaataaggca aacgtgtgtt gggccaaggc tttagtgcg	3120
gtgctgaaga ccgctggcat agacatgacc actgaacaat ggaacactgt ggattatatt	3180
gaaacggaca aagctcactc agcagagata gtattgaacc aactatgcgt gaggttcttt	3240
ggactcgatc tggactccgg tctatcttct gcacccactg ttccgttatc cattaggaat	3300
aatcactggg ataactcccc gtcgcctaac atgtacgggc tgaataaaga agtgggtccgt	3360
cagctctctc gcaggtaccc acaactgcct cgggcagttg cactggaag agtctatgac	3420
atgaacactg gtacactgcg caattatgat ccgcgcataa acctagtacc tgtaaacaga	3480
agactgcctc atgctttagt cctccaccat aatgaacacc cacagagtga cttttcttca	3540
ttcgtcagca aattgaaggc cagaactgtc ctgggtggtcg gggaaaagtt gtccgtccca	3600
ggcaaaatgg ttgactggtt gtcagaccgg cctgaggcta ctttcagagc tcggctggat	3660
ttaggcatcc caggtgatgt gcccaaatat gacataatat ttgttaatgt gaggacccca	3720

tataaataacc atcactatca gcagtgtgaa gaccatgcc a ttaagcttag catgttgacc	3780
aagaaagctt gtctgcatct gaatcccggc ggaacctgtg tcagcatagg ttatggttac	3840
gctgacaggg ccagcgaaag catcattggt gctatagcgc ggcagttcaa gttttcccgg	3900
gtatgcaaac cgaaatcctc acttgaagag acggaagttc tgtttgtatt cattgggtac	3960
gatcgcaagg cccgtacgca caatccttac aagctttcat caaccttgac caacatttat	4020
acaggttcca gactccacga agccggatgt gcaccctcat atcatgtggt gcgaggggat	4080
attgccacgg ccaccgaagg agtgattata aatgctgcta acagcaaagg acaacctggc	4140
ggaggggtgt gcggagcgct gtataagaaa ttcccggaaa gcttcgattt acagccgatc	4200
gaagtaggaa aagcgcgact ggtcaaaggt gcagctaaac atatcattca tgccgtagga	4260
ccaaacttca acaaagtttc ggaggttgaa ggtgacaaac agttggcaga ggcttatgag	4320
tccatcgcta agattgtcaa cgataacaat tacaagtcag tagcgattcc actgttgtcc	4380
accggcatct tttccgggaa caaagatcga ctaacccaat cattgaacca tttgctgaca	4440
gctttagaca ccactgatgc agatgtagcc atatactgca gggacaagaa atgggaaatg	4500
actctcaagg aagcagtggc taggagagaa gcagtggagg agatatgcat atccgacgac	4560
tcttcagtga cagaacctga tgcagagctg gtgagggtgc atccgaagag ttctttggct	4620
ggaaggaagg gctacagcac aagcgatggc aaaactttct catatttgga agggaccaag	4680
tttcaccagg cggccaagga tatagcagaa attaatgcc a tgtggcccgt tgcaacggag	4740
gccaatgagc aggtatgcat gtatatcctc ggagaaagca tgagcagtat taggtcgaag	4800
tgccccgtcg aagagtcgga agcctccaca ccacctagca cgctgccttg cttgtgcatc	4860
catgccatga ctccagaaag agtacagcgc ctaaaagcct cacgtccaga acaaattact	4920
gtgtgctcat cttttccatt gccgaagtat agaactactg gtgtgcagaa gatccaatgc	4980
tcccagccta tattgttctc accgaaagtg cctgcgtata ttcattccaag gaagtatctc	5040
gtggaaacac caccggtaga cgagactccg gagccatcgg cagagaacca atccacagag	5100
gggacacctg aacaaccacc acttataacc gaggatgaga ccaggactag aacgcctgag	5160
ccgatcatca tcgaagagga agaagaggat agcataagtt tgctgtcaga tggcccgacc	5220
caccaggtgc tgcaagtcga ggcagacatt cacgggccgc cctctgtatc tagctcatcc	5280
tgggtccattc ctcatgcatc cgactttgat gtggacagtt tatccatact tgacaccctg	5340

gagggagcta gcgtgaccag cggggcaacg tcagccgaga ctaactctta cttcgcaaag	5400
agtatggagt ttctggcgcg accggtgcct gcgcctcgaa cagtattcag gaaccctcca	5460
catcccgtc cgcgcacaaag aacaccgtca cttgcaccca gcagggcctg ctcgagaacc	5520
agcctagttt ccaccccgcc aggcgtgaat aggggtgatca ctagagagga gctcgaggcg	5580
cttaccccgt cacgcactcc tagcaggctg gtctcgagaa ccagcctggg ctccaacccg	5640
ccaggcgtaa atagggtgat tacaagagag gagtttgagg cgttcgtagc acaacaacaa	5700
tgacggtttg atgcgggtgc atacatcttt tcctccgaca ccggtcaagg gcatttacia	5760
caaaaatcag taaggcaaac ggtgctatcc gaagtgggtg tggagaggac cgaattggag	5820
atttcgtatg ccccgcgctt cgaccaagaa aaagaagaat tactacgcaa gaaattacag	5880
ttaaatccca cacctgctaa cagaagcaga taccagtcca ggaagggtga gaacatgaaa	5940
gccataacag ctagacgtat tctgcaaggc ctagggcatt atttgaaggc agaaggaaaa	6000
gtggagtgct accgaaccct gcctcctgtt cttttgtatt catctagtgt gaaccgtgcc	6060
ttttcaagcc ccaaggctgc agtggaagcc tgtaacgcca tggtgaaaga gaactttccg	6120
actgtggctt cttactgtat tattccagag tacgatgcct atttggacat ggttgacgga	6180
gcttcatgct gcttagacac tgccagtttt tgccctgcaa agctgcgcag ctttccaaag	6240
aaacactcct atttgggaacc cacaatacga tcggcagtgct cttcagcgat ccagaacacg	6300
ctccagaacg tcctggcagc tgccacaaaa agaaattgca atgtcacgca aatgagagaa	6360
ttgcccgtat tggattcggc ggcctttaat gtggaatgct tcaagaaata tgcgtgtaat	6420
aatgaatatt gggaaacgtt taaagaaaac cccatcaggc ttactgaaga aaacgtggta	6480
aattacatta ccaaattaaa aggacaaaaa gctgctgctc tttttgcgaa gacacataat	6540
ttgaatatgt tgcaggacat accaatggac aggtttgtaa tggacttaaa gagagacgtg	6600
aaagtgactc caggaacaaa acatactgaa gaacggccca aggtacagggt gatccaggct	6660
gccgatccgc tagcaacagc gtatctgtgc ggaatccacc gagagctggg taggagatta	6720
aatgcgggtc tgcttccgaa cattcataca ctgtttgata tgtcggctga agactttgac	6780
gctattatag ccgagcactt ccagcctggg gattgtgttc tggaaactga catcgcgctg	6840
tttgataaaa gtgaggacga cgccatggct ctgaccgcgt taatgattct ggaagactta	6900
ggtgtggacg cagagctgtt gacgctgatt gaggcggctt tcggcgaaat ttcacataa	6960

catttgccca ctaaaactaa atttaaattc ggagccatga tgaaatctgg aatgttcctc	7020
acactgtttg tgaacacagt cattaacatt gtaatcgcaa gcagagtgtt gagagaacgg	7080
ctaaccggat caccatgtgc agcattcatt ggagatgaca atatcgtgaa aggagtcaaa	7140
tcggacaaat taatggcaga caggtgcgcc acctggttga atatggaagt caagattata	7200
gatgctgtgg tgggcgagaa agcgccttat ttctgtggag ggtttatatt gtgtgactcc	7260
gtgaccggca cagcgtgccg tgtggcagac cccctaataa ggctgtttaa gcttggcaaa	7320
cctctggcag cagacgatga acatgatgat gacaggagaa gggcattgca tgaagagtca	7380
acacgctgga accgagtggg tattctttca gagctgtgca aggcagtaga atcaaggtat	7440
gaaaccgtag gaacttccat catagttatg gccatgacta ctctagctag cagtgttaaa	7500
tcattcagct acctgagagg ggcccctata actctctacg gctaacctga atggactacg	7560
actatcacgc ccaaacattt acagccgcgg tgtcaaaaac cgcgtggacg tggttaacat	7620
ccctgctggg aggatcagcc gtaattatta taattggctt ggtgctggct actattgtgg	7680
ccatgtacgt gctgaccaac cagaaacata attgaataca gcagcaattg gcaagctgct	7740
tacatagaac tcgcggcgat tggcatgccg ccttaaaatt tttatatttat tttttctttt	7800
cttttccgaa tcggattttg tttttaatat ttcaaaaaaa aaaaaaaaaa aaaaaaaaaa	7860
aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaatacgtag	7920
tttaaac	7927

<210> 9

<211> 7926

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic
polynucleotide

<400> 9

taatacgact cactatagga taggcggcgc atgagagaag cccagaccaa ttacctacc	60
aaaatggaga aagttcacgt tgacatcgag gaagacagcc cattcctcag agctttgcag	120
cggagcttcc cgcagtttga ggtagaagcc aagcagggtca ctgataatga ccatgcta	180
gccagagcgt tttcgcattt ggcttcaaaa ctgatcgaaa cggagggtgga cccatccgac	240

acgacacctg acattggaag tgcgcccgcc cgcagaatgt attctaagca caagtatcat	300
tgtatctgtc cgatgagatg tgcggaagat ccggacagat tgtataagta tgcaactaag	360
ctgaagaaaa actgtaagga aataactgat aaggaattgg acaagaaaat gaaggagctc	420
gccgccgtca tgagcgaccc tgacctggaa actgagacta tgtgcctcca cgacgacgag	480
tcgtgtcgct acgaagggca agtcgctgtt taccaggatg tatacgcggt tgacggaccg	540
acaagtctct atcaccaagc caataaggga gttagagtcg cctactggat aggctttgac	600
accacccctt ttatgtttta gaacttggct ggagcatatc catcatactc taccaactgg	660
gccgacgaaa ccgtgttaac ggctcgtaac ataggcctat gcagctctga cgttatggag	720
cggtcacgta gagggatgtc cattcttaga aagaagtatt tgaaaccatc caacaatgtt	780
ctattctctg ttggctcgac catctaccac gagaagaggg acttactgag gagctggcac	840
ctgccgtctg tatttcactt acgtggcaag caaaattaca catgtcgggtg tgagactata	900
gtagttgctg acgggtacgt cgttaaaaga atagctatca gtccaggcct gtatgggaag	960
ccttcaggct atgctgctac gatgcaccgc gagggattct tgtgctgcaa agtgacagac	1020
acattgaacg gggagagggt ctcttttccc gtgtgcacgt atgtgccagc tacattgtgt	1080
gaccaaata ga ctggcatact ggcaacagat gtcagtgcgg acgacgcgca aaaactgctg	1140
gttgggctca accagcgtat agtcgtcaac ggtcgcaccc agagaaacac caataccatg	1200
aaaaattacc ttttgcccgt agtggcccag gcatttgcta ggtgggcaaa ggaatataag	1260
gaagatcaag aagatgaaag gccactagga ctacgagata gacagttagt catgggggtgt	1320
tgttgggctt ttagaaggca caagataaca tctatttata agcgcccgga tacccaaacc	1380
atcatcaaag tgaacagcga tttccactca ttcgtgctgc ccaggatagg cagtaacaca	1440
ttggagatcg ggctgagaac aagaatcagg aaaatgttag aggagcacao ggagccgtca	1500
cctctcatta ccgccgagga cgtacaagaa gctaagtgcg cagccgatga ggctaaggag	1560
gtgcgtgaag ccgaggagtt gcgcgcagct ctaccacctt tggcagctga tgttgaggag	1620
cccactctgg aagccgatgt cgacttgatg ttacaagagg ctggggccgg ctcagtggag	1680
acacctctg gcttgataaa gggtaccagc tacgatggcg aggacaagat cggctcttac	1740
gctgtgcttt ctccgcaggc tgtactcaag agtgaaaaat tatcttgcac ccaccctctc	1800
gctgaacaag tcatagtgat aacacactct ggccgaaaag ggcgttatgc cgtggaacca	1860

taccatggta aagtagtggt gccagaggga catgcaatac ccgccagga ctttcaagct	1920
ctgagtga aa gtgccacat tgtgtacaac gaacgtgagt tcgtaaacag gtacctgcac	1980
catattgcca cacatggagg agcgctgaac actgatgaag aatattacaa aactgtcaag	2040
cccagcgagc acgacggcga atacctgtac gacatcgaca ggaaacagtg cgtcaagaaa	2100
gaactagtca ctgggctagg gctcacaggc gagctgggtg atcctccctt ccatgaattc	2160
gcctacgaga gtctgagaac acgaccagcc gctccttacc aagtaccaac cataggggtg	2220
tatggcgtgc caggatcagg caagtctggc atcattaaaa gcgcagtcac caaaaaagat	2280
ctagtgggtga gcgccaagaa agaaaactgt gcagaaatta taagggacgt caagaaaatg	2340
aaagggtg acgtcaatgc cagaactgtg gactcagtc tcttgaatgg atgcaaacac	2400
cccgtagaga ccctgtatat tgacgaagct tttgcttgtc atgcaggtac tctcagagcg	2460
ctcatagcca ttataagacc taaaaaggca gtgctctgcg gggatcccaa acagtgcggt	2520
ttttttaaca tgatgtgcct gaaagtgc at ttaaccacg agatttgcac acaagtcttc	2580
cacaaaagca tctctcgccg ttgcactaaa tctgtgactt cggtcgtctc aaccttgttt	2640
tacgacaaaa aaatgagaac gacgaatccg aaagagacta agattgtgat tgacactacc	2700
ggcagtagca aacctaagca ggacgatctc attctcactt gtttcagagg gtgggtgaag	2760
cagttgcaaa tagattacaa aggcaacgaa ataatgacgg cagctgcctc tcaagggtg	2820
acccgtaaaag gtgtgtatgc cgttcggtac aagggtgaatg aaaatcctct gtacgcaccc	2880
acctcagaac atgtgaacgt cctactgacc cgcacggagg accgcatcgt gtggaaaaca	2940
ctagccggcg acccatggat aaaaacactg actgccaagt accctgggaa tttcactgcc	3000
acgatagagg agtggcaagc agagcatgat gccatcatga ggcacatctt ggagagaccg	3060
gaccctaccg acgtcttcca gaataaggca aacgtgtgtt gggccaaggc tttagtgccg	3120
gtgctgaaga ccgctggcat agacatgacc actgaacaat ggaacactgt ggattatatt	3180
gaaacggaca aagctcactc agcagagata gtattgaacc aactatgcgt gaggttcttt	3240
ggactcgtatc tggactccgg tctatcttct gcacccactg ttccgttatc cattaggaat	3300
aatcactggg ataactcccc gtcgcctaac atgtacgggc tgaataaaga agtgggtccgt	3360
cagctctctc gcaggtaccc acaactgcct cgggcagttg ccactggaag agtctatgac	3420
atgaacactg gtacactgcg caattatgat ccgcgcataa acctagtacc tgtaaacaga	3480

agactgcctc atgcttttagt cctccaccat aatgaacacc cacagagtga cttttcttca	3540
ttcgtcagca aattgaaggg cagaactgtc ctgggtggtcg gggaaaagtt gtccgtccca	3600
ggcaaaatgg ttgactgggtt gtcagaccgg cctgaggcta ccttcagagc tcggctggat	3660
ttaggcatcc caggtgatgt gcccaaatat gacataatat ttgttaatgt gaggaccca	3720
tataaatacc atcactatca gcagtgtgaa gaccatgcca ttaagcttag catgttgacc	3780
aagaaagctt gtctgcatct gaatcccggc ggaacctgtg tcagcatagg ttatggttac	3840
gctgacaggg ccagcgaaag catcattgggt gctatagcgc ggcagttcaa gttttcccgg	3900
gtatgcaaac cgaaatcctc acttgaagag acggaagttc tgtttgtatt cattgggtac	3960
gatcgcaagg cccgtacgca caatccttac aagctttcat caaccttgac caacatttat	4020
acaggttcca gactccacga agccggatgt gcaccctcat atcatgtggt gcgaggggat	4080
attgccacgg ccaccgaagg agtgattata aatgctgcta acagcaaagg acaacctggc	4140
ggaggggtgt gcggagcgct gtataagaaa ttcccggaaa gcttcgattt acagccgatc	4200
gaagtaggaa aagcgcgact ggtcaaaggt gcagctaaac atatcattca tgccgtagga	4260
ccaaacttca acaaagtttc ggaggttgaa ggtgacaaac agttggcaga ggcttatgag	4320
tccatcgcta agattgtcaa cgataacaat tacaagtcag tagcgattcc actgttgtcc	4380
accggcatct tttccgggaa caaagatcga ctaacccaat cattgaacca tttgctgaca	4440
gcttttagaca ccactgatgc agatgtagcc atatactgca gggacaagaa atgggaaatg	4500
actctcaagg aagcagtggc taggagagaa gcagtggagg agatatgcat atccgacgac	4560
tcttcagtga cagaacctga tgcagagctg gtgagggtgc atccgaagag ttctttggct	4620
ggaaggaagg gctacagcac aagcgatggc aaaactttct catatttgga agggaccaag	4680
tttcaccagg cggccaagga tatagcagaa attaatgcca tgtggcccgt tgcaacggag	4740
gccaatgagc aggtatgcat gtatatcctc ggagaaagca tgagcagtat taggtcga	4800
tgccccgtcg aagagtcgga agcctccaca ccacctagca cgctgccttg cttgtgcatc	4860
catgccatga ctccagaaag agtacagcgc ctaaaagcct cacgtccaga acaaattact	4920
gtgtgctcat cttttccatt gccgaagtat agaactactg gtgtgcagaa gatccaatgc	4980
tcccagccta tattgttctc accgaaagtg cctgcgtata ttcaccaag gaagtatctc	5040
gtggaaacac caccggtaga cgagactccg gagccatcgg cagagaacca atccacagag	5100

gggacacctg aacaaccacc atttataacc gaggatgaga ccaggactag aacgcctgag	5160
ccgatcatca tcgaagagga agaagaggat agcataagtt tgctgtcaga tggcccgacc	5220
caccaggtgc tgcaagtcga ggcagacatt cacgggccgc cctctgtatc tagctcatcc	5280
tgggtccattc ctcatgcatc cgactttgat gtggacagtt tatccatact tgacaccctg	5340
gagggagcta gcgtgaccag cggggcaacg tcagccgaga ctaactctta cttcgcaaag	5400
agtatggagt ttctggcgcg accggtgcct gcgcctcgaa cagtattcag gaaccctcca	5460
catcccgtc cgcgcacaaag aacaccgtca cttgcaccca gcagggcctg ctcgagaacc	5520
agcctagttt ccaccccgc aggcgtgaat aggggtgatca ctagagagga gctcgaggcg	5580
cttaccccgt cacgcactcc tagcaggtcg gtctcgagaa ccagcctggt ctccaacccg	5640
ccaggcgtaa atagggtgat tacaagagag gagtttgagg cgttcgtagc acaacaacaa	5700
tgacggtttg atgcgggtgc atacatcttt tcctccgaca ccggtcaagg gcatttacia	5760
caaaaatcag taaggcaaac ggtgctatcc gaagtgggtg tggagaggac cgaattggag	5820
atttcgtatg ccccgcgctt cgaccaagaa aaagaagaat tactacgcaa gaaattacag	5880
ttaaatccca cacctgctaa cagaagcaga taccagtcca ggaaggtgga gaacatgaaa	5940
gccataacag ctagacgtat tctgcaaggc ctagggcatt atttgaaggc agaaggaaaa	6000
gtggagtgct accgaaccct gcatcctggt cttttgtatt catctagtgt gaaccgtgcc	6060
ttttcaagcc ccaaggtcgc agtggaagcc tgtaacgcca tgttgaaaga gaactttccg	6120
actgtggctt cttactgtat tattccagag tacgatgcct atttggacat ggttgacgga	6180
gcttcatgct gcttagacac tgccagtttt tgccctgcaa agctgcgcag ctttccaaag	6240
aaacactcct atttgaacc cacaatacga tcggcagtgct cttcagcgat ccagaacacg	6300
ctccagaacg tcctggcagc tgccacaaaa agaaattgca atgtcacgca aatgagagaa	6360
ttgcccgtat tggattcggc ggcctttaat gtggaatgct tcaagaaata tgcgtgtaat	6420
aatgaatatt gggaaacgtt taaagaaaac cccatcaggc ttactgaaga aaacgtggta	6480
aattacatta ccaaattaaa aggaccaaaa gctgctgctc tttttgcgaa gacacataat	6540
ttgaatatgt tgcaggacat accaatggac aggtttgtaa tggacttaaa gagagacgtg	6600
aaagtgactc caggaacaaa acatactgaa gaacggccca aggtacaggt gatccaggct	6660
gccgatccgc tagcaacagc gtatctgtgc ggaatccacc gagagctggt taggagatta	6720

aatgcggtcc tgcttccgaa cattcataca ctgtttgata tgtcggctga agactttgac	6780
gctattatag ccgagcactt ccagcctggg gatttgtgtt tggaactga catcgcgtcg	6840
tttgataaaa gtgaggacga cgccatggct ctgaccgct taatgattct ggaagactta	6900
ggtgtggacg cagagctggt gacgctgatt gaggcggctt tcggcgaaat ttcataata	6960
catttgccca ctaaaactaa atttaaattc ggagccatga tgaaatctgg aatgttcctc	7020
acactgtttg tgaacacagt cattaacatt gtaatcgcaa gcagagtgtt gagagaacgg	7080
ctaaccggat caccatgtgc agcattcatt ggagatgaca atatcgtgaa aggagtcaaa	7140
tcggacaaat taatggcaga caggtgcgcc acctggttga atatggaagt caagattata	7200
gatgctgtgg tgggcgagaa agcgccttat ttctgtggag ggtttatttt gtgtgactcc	7260
gtgaccggca cagcgtgccg tgtggcagac cccctaaaaa ggctgtttaa gcttggcaaa	7320
cctctggcag cagacgatga acatgatgat gacaggagaa gggcattgca tgaagagtca	7380
acacgctgga accgagtggg tattctttca gagctgtgca aggcagtaga atcaaggtat	7440
gaaaccgtag gaacttccat catagttagt gccatgacta ctctagctag cagtgttaaa	7500
tcattcagct acctgagagg ggcccctata actctctacg gctaacctga atggactacg	7560
actatcacgc ccaaacatth acagccgcgg tgtcaaaaac cgcgtggacg tggttaacat	7620
ccctgctggg aggatcagcc gtaattatta taattggctt ggtgctggct actattgtgg	7680
ccatgtacgt gctgaccaac cagaaacata attgaataca gcagcaattg gcaagctgct	7740
tacatagaac tcgcggcgat tggcatgccg ccttaaaatt tttattttat ttttcttttc	7800
ttttccgaat cggattttgt ttttaatat tcaaaaaaaaa aaaaaaaaaa aaaaaaaaaa	7860
aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aatacgtagt	7920
ttaaac	7926

<210> 10

<211> 36519

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 10

ccatcttcaa taatatacct caaacttttt gtgcgcgtta atatgcaaat gaggcgtttg	60

aatttgggga ggaagggcgg tgattggtcg agggatgagc gaccgttagg ggcggggcga	120
gtgacgtttt gatgacgtgg ttgcgaggag gagccagttt gcaagttctc gtgggaaaag	180
tgacgtcaaa cgaggtgtgg tttgaacacg gaaataactca attttcccgc gctctctgac	240
aggaaatgag gtgttttctgg gcggatgcaa gtgaaaacgg gccatttttcg cgcgaaaact	300
gaatgaggaa gtgaaaatct gagtaatttc gcgtttatgg caggaggagg tatttgccga	360
gggccgagta gactttgacc gattacgtgg gggtttcgat taccgtgttt ttcacctaaa	420
tttccgcgta cgggtgtcaaa gtccggtgtt tttacgtagg tgtcagctga tcgccagggt	480
atttaaact gcgctctcca gtcaagaggc cactcttgag tgccagcgag aagagttttc	540
tcctccgcgc cgcgagtcag atctacactt tgaaagatga ggcacctgag agacctgccc	600
gatgagaaaa tcatcatcgc ttccgggaac gagattctgg aactgggtgt aaatgccatg	660
atgggcgacg accctccgga gccccccacc ccatttgaga caccttcgct gcacgatttg	720
tatgatctgg aggtggatgt gcccaggagc gatcccaatg aggaggcgggt aaatgatttt	780
tttagcgatg ccgcgctgct agctgccgag gaggcttcga gctctagctc agacagcgac	840
tcttactgc ataccctag acccggcaga ggtgagaaaa agatccccga gcttaaaggg	900
gaagagatgg acttgcgctg ctatgaggaa tgcttgcccc cgagcgatga tgaggacgag	960
caggcgatcc agaacgcagc gagccaggga gtgcaagccg ccagcgagag ctttgcgctg	1020
gactgcccgc ctctgcccgg acacggctgt aagtcttggtg aatttcacgc catgaatact	1080
ggagataaag ctgtgttggtg tgcactttgc tatatgagag cttacaacca ttgtgtttac	1140
agtaagtgtg attaagttga acttttagagg gaggcagaga gcagggtgac tgggcgatga	1200
ctggtttatt tatgtatata tgttctttat ataggtcccg tctctgacgc agatgatgag	1260
acccccacta caaagtccac ttcgtcacc ccagaaattg gcacatctcc acctgagaat	1320
attgttagac cagttcctgt tagagccact gggaggagag cagctgtgga atgtttggat	1380
gacttgctac aggggtgggg tgaacctttg gacttggtgta cccggaaacg ccccaggcac	1440
taagtgccac acatgtgtgt ttacttgagg tgatgtcagt atttataggg tgtggagtgc	1500
aataaaaaat gtgttgactt taagtgcgtg gtttatgact caggggtggg gactgtgagt	1560
atataagcag gtgcagacct gtgtggttag ctcagagcgg catggagatt tggacggtct	1620
tggaagactt tcacaagact agacagctgc tagagaacgc ctcgaacgga gtctcttacc	1680

tgtggagatt	ctgcttcggt	ggcgacctag	ctaggctagt	ctacagggcc	aaacaggatt	1740
atagtgaaca	atttgagggt	attttgagag	agtgttctgg	tctttttgac	gctcttaact	1800
tgggccatca	gtctcacttt	aaccagagga	tttcgagagc	ccttgatttt	actactcctg	1860
gcagaaccac	tgcagcagta	gccttttttg	cttttattct	tgacaaatgg	agtcaagaaa	1920
cccatttcag	cagggattac	cagctggatt	tcttagcagt	agctttgtgg	agaacatgga	1980
agtgccagcg	cctgaatgca	atctccggct	acttgccggg	acagccgcta	gacactctga	2040
ggatcctgaa	tctccaggag	agtcccaggg	cacgccaacg	tcgccagcag	cagcagcagg	2100
aggaggatca	agaagagaac	ccgagagccg	gcctggaccc	tccggcggag	gaggaggagt	2160
agctgacctg	tttctgaac	tgcgccgggt	gctgactagg	tcttcgagtg	gtcgggagag	2220
ggggattaag	cgggagaggc	atgatgagac	taatcacaga	actgaactga	ctgtgggtct	2280
gatgagtcgc	aagcgcccag	aaacagtgtg	gtggcatgag	gtgcagtcga	ctggcacaga	2340
tgaggtgtcg	gtgatgcatg	agaggttttc	tctagaacaa	gtcaagactt	gttggttaga	2400
gcctgaggat	gattgggagg	tagccatcag	gaattatgcc	aagctggctc	tgaggccaga	2460
caagaagtac	aagattacta	agctgataaa	tatcagaaat	gcctgctaca	tctcagggaa	2520
tggggctgaa	gtggagatct	gtctccagga	aagggtggct	ttcagatgct	gcatgatgaa	2580
tatgtacccg	ggagtgggtg	gcatggatgg	ggttaccttt	atgaacatga	ggttcagggg	2640
agatgggtat	aatggcacgg	tctttatggc	caataccaag	ctgacagtcc	atggctgctc	2700
cttctttggg	tttaataaca	cctgcatcga	ggcctggggg	caggtcgggtg	tgaggggctg	2760
cagtttttca	gccaaactgga	tgggggtcgt	gggcaggacc	aagagtatgc	tgtccgtgaa	2820
gaaatgcttg	tttgagaggt	gccacctggg	ggtgatgagc	gagggcgaag	ccagaatccg	2880
ccactgcgcc	tctaccgaga	cgggctgctt	tgtgctgtgc	aagggaatg	ctaagatcaa	2940
gcataatatg	atctgtggag	cctcggacga	gcgcggctac	cagatgctga	cctgcgccgg	3000
cgggaacagc	catatgctgg	ccaccgtaca	tgtggcttcc	catgctcgca	agccctggcc	3060
cgagttcgag	cacaatgtca	tgaccaggtg	caatatgcat	ctggggctcc	gccgaggcat	3120
gttcatgccc	taccagtgca	acctgaatta	tgtgaagggtg	ctgctggagc	ccgatgccat	3180
gtccagagtg	agcctgacgg	gggtgtttga	catgaatgtg	gaggtgtgga	agattctgag	3240
atatgatgaa	tccaagacca	ggtgccgagc	ctgcgagtg	ggagggaagc	atgccaggtt	3300

ccagcccgtg tgtgtggatg tgacggagga cctgcgaccc gatcatttgg tgttgccctg	3360
caccggggacg gagttcggtt ccagcgggga agaatctgac tagagtgagt agtgttctgg	3420
ggcgggggag gacctgcatg agggccagaa taactgaaat ctgtgctttt ctgtgtgttg	3480
cagcagcatg agcggaagcg gctcctttga gggaggggta ttcagccctt atctgacggg	3540
gcgtctcccc tcctgggcgg gagtgcgta gaatgtgat ggatccacgg tggacggccg	3600
gcccgtgcag cccgcgaact cttcaaccct gacctatgca accctgagct cttcgtcgtt	3660
ggacgcagct gccgccgag ctgctgcatc tgccgccagc gccgtgcgcg gaatggccat	3720
gggcgccggc tactacggca ctctggtggc caactcgagt tccaccaata atccccccag	3780
cctgaacgag gagaagctgt tgctgctgat ggcccagctc gaggccttga cccagcgcct	3840
gggcgagctg acccagcagg tggctcagct gcaggagcag acgcgggccg cggttgccac	3900
ggtgaaatcc aaataaaaaa tgaatcaata aataaacgga gacggttggt gattttaaca	3960
cagagtctga atctttatctt gatctttcgc gcgcggtagg ccctggacca ccggtctcga	4020
tcattgagca cccggtggat cttttccagg acccggtaga ggtgggcttg gatgttgagg	4080
tacatgggca tgagcccgtc ccggggggtg aggtagctcc attgcagggc ctcgtgctcg	4140
ggggtggtgt tgtaaatac ccagtcatag caggggcgca gggcatggtg ttgcacaata	4200
tctttgagga ggagactgat ggccacgggc agccctttgg tgtaggtgtt taaaaatctg	4260
ttgagctggg agggatgcat gcgggggggag atgagggtgca tcttggcctg gatcttgaga	4320
ttggcgatgt taccgcccag atccgcctg gggttcatgt tgtgcaggac caccagcacg	4380
gtgtatccgg tgcacttggg gaatttatca tgcaacttgg aagggaaggc gtgaaagaat	4440
ttggcgacgc ctttgtgccc gccaggttt tccatgcact catccatgat gatggcgatg	4500
ggcccgtggg cggcggcctg ggcaaagacg tttcgggggt cggacacatc atagtgttg	4560
tcctgggtga ggtcatcata ggccatttta atgaatttgg ggcggagggt gccggactgg	4620
gggacaaagg taccctcgat cccgggggcg tagttcccct cacagatctg catctcccag	4680
gctttgagct cggagggggg gatcatgtcc acctgcgggg cgataaagaa cacggtttcc	4740
ggggcggggg agatgagctg ggccgaaagc aagttccgga gcagctggga cttgccgcag	4800
ccggtggggc cgtagatgac cccgatgacc ggctgcaggt ggtagttgag ggagagacag	4860
ctgccgtcct cccggaggag gggggccacc tcgttcatca tctcgcgcac gtgcatgttc	4920

tcgcgcacca gttccgccag gaggcgctct cccccaggg ataggagctc ctggagcgag	4980
gcgaagtttt tcagcggctt gagtccgtcg gccatgggca ttttggagag ggtttgttgc	5040
aagagttcca ggcggtccca gagctcggtg atgtgctcta cggcatctcg atccagcaga	5100
cctcctcgtt tcgcgggttg ggacggctgc gggagtaggg caccagacga tgggcgtcca	5160
gcgcagccag ggtccggtcc ttccagggtc gcagcgtccg cgtcagggtg gtctccgtca	5220
cggngaaggg gtgcgcgccg ggctgggcgc ttgcgagggt gcgcttcagg ctcatccggc	5280
tggtcgaaaa ccgctcccga tcggcgcctt gcgcgtcggc caggtagcaa ttgacatga	5340
gttcgtagtt gagcgctcg gccgcgtggc ctttggcgcg gagcttacct ttggaagtct	5400
gcccgcaggc gggacagagg agggacttga gggcgtagag cttgggggcg aggaagacgg	5460
actcgggggc gtaggcgtcc gcgccgcagt gggcgcagac ggtctcgcac tccacgagcc	5520
aggtgaggtc gggctggtcg ggggtcaaaaa ccagtttccc gccgttcttt ttgatgcgtt	5580
tcttaccttt ggtctccatg agctcgtgtc cccgctgggt gacaaagagg ctgtccgtgt	5640
ccccgtagac cgactttatg ggccggtcct cgagcgggtg gccgcgggtc tcctcgtaga	5700
ggaacccgc ccactccgag acgaaagccc ggggtccaggc cagcacgaag gaggccacgt	5760
gggacgggta gcggtcgttg tccaccagcg ggtccacctt ttccagggta tgcaaacaca	5820
tgtccccctc gtccacatcc aggaaggtga ttggcttgta agtgtaggcc acgtgaccgg	5880
gggtcccggc cgggggggta taaaagggtg cgggtccctg ctcgtcctca ctgtcttccg	5940
gatcgtgtc caggagcgcc agctgttggg gtaggtattc cctctcgaag gcgggcatga	6000
cctcggcact caggttgtca gtttctagaa acgaggagga tttgatattg acggtgccgg	6060
cggagatgcc tttcaagagc ccctcgtcca tctggtcaga aaagacgatc tttttgttgt	6120
cgagcttggg ggcgaaggag ccgtagaggg cgttggagag gagcttggcg atggagcgca	6180
tggtctggtt tttttccttg tcggcgcgct ctttggcggc gatgttgagc tgcacgtact	6240
cgcgcgccac gcacttccat tcggggaaga cggtggtcag ctcgtcgggc acgattctga	6300
cctgccagcc ccgattatgc agggatga ggtccacact ggtggccacc tcgccgcgca	6360
ggggctcatt agtccagcag aggcgtccgc ctttgcgcga gcagaagggg ggcaggggggt	6420
ccagcatgac ctcgtcgggg gggtcggcat cgatggtgaa gatgccgggc aggaggtcgg	6480
ggtcaaagta gctgatggaa gtggccagat cgtccagggc agcttgccat tcgcgcacgg	6540

ccagcgcgcg	ctcgtagggg	ctgagggg	tgccccaggg	catgggatgg	gtaagcgcgg	6600
aggcgtacat	gccgcagatg	tcgtagacgt	agaggggctc	ctcgaggatg	ccgatgtagg	6660
tggggtagca	gcgccccccg	cggatgctgg	cgcgcacgta	gtcatacagc	tcgtgcgagg	6720
gggcgaggag	ccccggggcc	aggttggtgc	gactgggctt	ttcggcgcgg	tagacgatct	6780
ggcggaaaat	ggcatgcgag	ttggaggaga	tggtgggcct	ttggaagatg	ttgaagtggg	6840
cgtggggcag	tccgaccgag	tcgcggatga	agtgggcgta	ggagtcttgc	agcttggcga	6900
cgagctcggc	ggtgactagg	acgtccagag	cgcagtagtc	gagggtctcc	tggatgatgt	6960
catacttgag	ctgtcccttt	tgtttccaca	gctcgcggtt	gagaaggaac	tcttcgcggt	7020
ccttccagta	ctcttcgagg	gggaacccgt	cctgatctgc	acggtaagag	cctagcatgt	7080
agaactgggt	gacggccttg	taggcgcagc	agcccttctc	cacggggagg	gcgtaggcct	7140
gggcggcctt	gcgcagggag	gtgtgcgtga	gggcgaaagt	gtccctgacc	atgaccttga	7200
ggaactggtg	cttgaagtcg	atatcgtcgc	agccccctg	ctcccagagc	tggaagtccg	7260
tgcgcttctt	gtaggcgggg	ttgggcaaag	cgaaagtaac	atcgttgaag	aggatcttgc	7320
ccgcgcgggg	cataaagttg	cgagtgatgc	ggaaaggttg	gggcacctcg	gcccggttgt	7380
tgatgacctg	ggcggcgagc	acgatctcgt	cgaagccgtt	gatgttgtgg	cccacgatgt	7440
agagtccac	gaatcgcgga	cggcccttga	cgtggggcag	tttcttgagc	tcctcgtagg	7500
tgagctcgtc	ggggtcgctg	agcccgtgct	gctcgagcgc	ccagtcggcg	agatgggggt	7560
tggcgcggag	gaaggaagtc	cagagatcca	cggccagggc	ggtttgaga	cgggtcccgt	7620
actgacggaa	ctgctgcccg	acggccattt	tttcgggggt	gacgcagtag	aaggatgcggg	7680
ggtccccgtg	ccagcgatcc	catttgagct	ggagggcgag	atcgagggcg	agctcgacga	7740
gccggtcgtc	cccggagagt	ttcatgacca	gcatgaaggg	gacgagctgc	ttgccgaagg	7800
accccatcca	ggtgtagggt	tccacatcgt	aggtgaggaa	gagcctttcg	gtgcgaggat	7860
gcgagccgat	ggggaagaac	tgatctcct	gccaccaatt	ggaggaatgg	ctgttgatgt	7920
gatggaagta	gaaatgccga	cggcgcgccg	aacactcgtg	cttgtgttta	tacaagcggc	7980
cacagtgtc	gcaacgtgc	acgggatgca	cgtgtgcac	gagctgtacc	tgagttcctt	8040
tgacgaggaa	tttcagtggg	aagtggagtc	gtggcgcctg	catctcgtgc	tgtactacgt	8100
cgtgggtggtc	ggcctggccc	tcttctgcct	cgatgggtgg	catgctgacg	agcccgcgcg	8160

ggaggcaggt ccagacctcg gcgcgagcgg gtcggagagc gaggacgagg gcgcgcaggc	8220
cggagctgtc cagggtcctg agacgctgcg gagtcaaggc agtgggcagc ggccggcgcg	8280
ggttgacttg caggagtttt tccagggcg cggggagggtc cagatggtag ttgatctcca	8340
ccgcgccatt ggtggcgacg tcgatggctt gcaggggtccc gtgcccctgg ggtgtgacca	8400
ccgtcccccg tttcttcttg ggccggctggg gcgacggggg cggtagcctt tccatggtta	8460
gaagcggcgg cgaggacgcg cgccgggcgg caggggcggc tcggggcccg gaggcagggg	8520
cggcaggggc acgtcggcg cgcgcgcggg taggttcttg tactgcgccc ggagaagact	8580
ggcgtgagcg acgacgcgac ggttgacgtc ctggatctga cgctctggg tgaaggccac	8640
gggacccgtg agtttgaacc tgaaagagag ttgacagaa tcaatctcgg tatcgttgac	8700
ggcggcctgc cgcaggatct cttgcacgtc gcccgagttg tcctggtagg cgatctcgg	8760
catgaactgc tcgatctcct cctcttgaag gtctccgcgg ccggcgcgct ccacggtggc	8820
cgcgaggtcg ttggagatgc ggcccatgag ctgcgagaag gcgttcatgc ccgcctcgtt	8880
ccagacgcgg ctgtagacca cgacgccctc gggatcgcg gcgcgcatga ccacctgggc	8940
gaggttgagc tccacgtggc gcgtgaagac cgcgtagttg cagaggcgct ggtagaggta	9000
gttgagcgtg gtggcgatgt gctcggtagc gaagaaatac atgatccagc ggccggagcg	9060
catctcgtg acgtcgccca gcgcctcaa acgttccatg gcctcgtaaa agtccacggc	9120
gaagttgaaa aactgggagt tgccgcgccga gacggtcaac tcctcctcca gaagacggat	9180
gagctcggcg atggtggcg gcacctcgcg ctggaaggcc cccgggagtt cctccacttc	9240
ctcttcttcc tcctccacta acatctcttc tacttcttcc tcaggcggca gtggtggcg	9300
gggagggggc ctgcgtcgcc ggccggcgac gggcagacgg tcgatgaagc gctcgatggt	9360
ctgccgcgc cggcgtcgca tgggtctcgg gacggcgcg ccgtcctcg ggggccgcag	9420
cgtgaagacg ccgccgcga tctccagggt gccggggggg tccccgttgg gcaggagag	9480
ggcgctgacg atgcatctta tcaattgccc cgtagggact ccgcgcaagg acctgagcgt	9540
ctcgagatcc acgggatctg aaaaccgtg aacgaaggct tcgagccagt cgagtcgca	9600
aggtagggtg agcacggttt cttctggcgg gtcagtgttg ttgggagcgg ggccggcgat	9660
gctgctggtg atgaagttga aataggcggt tctgagacgg cggtatggtg cgaggagcac	9720
caggtctttg ggcccggctt gctggatgcg cagacggtcg gccatgcccc aggcgtggtc	9780

ctgacacctg gccaggtcct tgtagtagtc ctgcatgagc cgctccacgg gcacctcctc	9840
ctcgcccgcg cggccgtgca tgcgcgtgag cccgaagccg cgctggggct ggacgagcgc	9900
caggtcggcg acgacgcgct cggcgaggat ggcttgctgg atctgggtga ggggtggtctg	9960
gaagtcatca aagtcgacga agcgggtgga ggctccggtg ttgatgggtg aggagcagtt	10020
ggccatgacg gaccagttga cggctctggtg gcccggacgc acgagctcgt ggtacttgag	10080
gcgcgagtag gcgcgcgtgt cgaagatgta gtcgttgacg gtgcgcacca ggtactggta	10140
gccgatgagg aagtgcggcg gcggctggcg gtagagcggc catcgctcgg tggcgggggc	10200
gccgggcgcg aggtcctcga gcatgggtgcg gtggtagccg tagatgtacc tggacatcca	10260
ggtgatgccg gcggcggtgg tggaggcgcg cgggaactcg cggacgcggt tccagatgtt	10320
gcgcagcggc aggaagtagt tcatggtggg cacggtctgg cccgtgaggc gcgcgcagtc	10380
gtggatgctc tatacgggca aaaacgaaag cggtcagcgg ctcgactccg tggcctggag	10440
gctaagcgaa cgggttgggc tgcgcgtgta ccccggttcg aatctcgaat caggctggag	10500
ccgcagctaa cgtgggtattg gcactcccgt ctcgacccaa gcctgcacca accctccagg	10560
atacggaggc gggtcgtttt gcaacttttt tttggaggcc ggatgagact agtaagcgcg	10620
gaaagcggcc gaccgcgatg gctcgtgcc gtagtctgga gaagaatcgc cagggttgcg	10680
ttgcggtgtg ccccggttcg aggccggccg gattccgcgg ctaacgaggg cgtggctgcc	10740
ccgtcgtttc caagacccca tagccagccg acttctccag ttacggagcg agcccctctt	10800
ttgttttgtt tgtttttgcc agatgcatcc cgtactgcgg cagatgcgcc cccaccacc	10860
tccaccgcaa caacagcccc ctccacagcc ggcgttctg ccccgcccc agcagcaact	10920
tccagccacg accgccgcgg ccgccgtgag cggggctgga cagagttatg atcaccagct	10980
ggccttgga gaggcgagg ggctggcgcg cctgggggcg tcgtcgccgg agcggcaccc	11040
gcgcgtgcag atgaaaagg acgctcgcga ggcctacgtg cccaagcaga acctgttcag	11100
agacaggagc ggcgaggagc ccgaggagat gcgcgcggcc cggttcacg cggggcgagg	11160
gctgcggcgc ggcctggacc gaaagagggt gctgagggac gaggatttcg aggcggacga	11220
gctgacgggg atcagccccg cgcgcgcgca cgtggccgcg gccaacctgg tcacggcgta	11280
cgagcagacc gtgaaggagg agagcaactt ccaaaaatcc ttcaacaacc acgtgcgcac	11340
cctgatcgcg cgcgaggagg tgaccctggg cctgatgcac ctgtgggacc tgctggaggc	11400

catcgtgcag aacccccacca gcaagccgct gacggcgcag ctgttcctgg tgggtgcagca	11460
tagtcgggac aacgaagcgt tcagggaggc gctgctgaat atcaccgagc ccgaggggccg	11520
ctggctcctg gacctggtga acattctgca gagcatcgtg gtgcaggagc gcgggctgcc	11580
gctgtccgag aagctggcgg ccatcaactt ctcggtgctg agtttgggca agtactacgc	11640
taggaagatc tacaagaccc cgtacgtgcc catagacaag gaggtgaaga tcgacggggtt	11700
ttacatgcgc atgaccctga aagtgtgac cctgagcgac gatctggggg tgtaccgcaa	11760
cgacaggatg caccgtgcgg tgagcgccag caggcggcgc gagctgagcg accaggagct	11820
gatgcatagt ctgcagcggg ccctgaccgg ggccggggacc gagggggaga gctactttga	11880
catgggcgcg gacctgcact ggcagcccag ccgccgggcc ttggaggcgg cggcaggacc	11940
ctacgtagaa gaggtggacg atgaggtgga cgaggagggc gagtacctgg aagactgatg	12000
gcgcgaccgt atttttgcta gatgcaacaa caacagccac ctcctgatcc cgcgatgcgg	12060
gcggcgctgc agagccagcc gtccggcatt aactcctcgg acgattggac ccaggccatg	12120
caacgcatca tggcgctgac gacccgcaac cccgaagcct ttagacagca gccccaggcc	12180
aaccggctct cggccatcct ggaggccgtg gtgccctcgc gctccaaccc cacgcacgag	12240
aaggtcctgg ccatcgtgaa cgcgctggtg gagaacaagg ccatccgcgg cgacgaggcc	12300
ggcctggtgt acaacgcgt gctggagcgc gtggcccgt acaacagcac caacgtgcag	12360
accaacctgg accgcatggt gaccgacgtg cgcgaggccg tggcccagcg cgagcggttc	12420
caccgcgagt ccaacctggg atccatggtg gcgctgaacg ccttcctcag caccagccc	12480
gccaacgtgc cccggggcca ggaggactac accaacttca tcagcgccct gcgcctgatg	12540
gtgaccgagg tgccccagag cgaggtgtac cagtccgggc cggactactt cttccagacc	12600
agtcgccagg gcttgacagc cgtgaacctg agccaggctt tcaagaactt gcagggcctg	12660
tggggcgctgc aggccccggt cggggaccgc gcgacggtgt cgagcctgct gacgccgaac	12720
tcgcgctgc tgctgctgct ggtggcccc ttcacggaca gcggcagcat caaccgcaac	12780
tcgtacctgg gctacctgat taacctgtac cgcgaggcca tcggccaggc gcacgtggac	12840
gagcagacct accaggagat caccacgtg agccgcgccc tgggccagga cgacccgggc	12900
aacctggaag ccaccctgaa ctttttgctg accaaccggt cgcagaagat cccgccccag	12960
tacgcgctca gcaccgagga ggagcgcac ctcggttacg tgcagcagag cgtgggcctg	13020

ttcctgatgc	aggagggggc	cacccccagc	gccgcgctcg	acatgaccgc	gcgcaacatg	13080
gagcccagca	tgtacgccag	caaccgcccc	ttcatcaata	aactgatgga	ctacttgcat	13140
cgggcggccg	ccatgaactc	tgactatttc	accaacgcca	tcctgaatcc	ccactggctc	13200
ccgccgccgg	ggttctacac	gggcgagtac	gacatgcccc	accccaatga	cgggttcctg	13260
tgggacgatg	tggacagcag	cgtgttctcc	ccccgaccgg	gtgctaacga	gcgccccttg	13320
tggaagaagg	aaggcagcga	ccgacgcccc	tcctcggcgc	tgtccggccg	cgagggtgct	13380
gccgcggcgg	tgcccagagg	cgccagtcct	ttcccagagct	tgcccttctc	gctgaacagt	13440
atccgcagca	gcgagctggg	caggatcacg	cggccgcgct	tgctgggcga	agaggagtac	13500
ttgaatgact	cgctgttgag	acccgagcgg	gagaagaact	tccccaataa	cgggatagaa	13560
agcctggtgg	acaagatgag	ccgctggaag	acgtatgcgc	aggagcacag	ggacgatccc	13620
cgggcgtcgc	agggggccac	gagccggggc	agcgcgccc	gtaaacgccg	gtggcacgac	13680
aggcagcggg	gacagatgtg	ggacgatgag	gactccgccg	acgacagcag	cgtgttggac	13740
ttgggtggga	gtggtaacct	gttcgctcac	ctgcgcccc	gtatcgggcg	catgatgtaa	13800
gagaaaccga	aaataaatga	tactcaccaa	ggccatggcg	accagcgtgc	gttcgtttct	13860
tctctgttgt	tgttgtatct	agtatgatga	ggcgtgcgta	cccggagggt	cctcctccct	13920
cgtacgagag	cgtgatgcag	caggcgatgg	cggcggcggc	gatgcagccc	ccgctggagg	13980
ctccttacgt	gccccgcgg	tacctggcgc	ctacggaggg	gcggaacagc	attcgttact	14040
cggagctggc	acccttgtac	gataccacct	ggttgtacct	ggtggacaac	aagtcggcgg	14100
acatgcctc	gctgaactac	cagaacgacc	acagcaactt	cctgaccacc	gtggtgcaga	14160
acaatgactt	cacccccacg	gaggccagca	cccagaccat	caactttgac	gagcgcctgc	14220
ggtggggcgg	ccagctgaaa	accatcatgc	acaccaacat	gcccacgtg	aacgagttca	14280
tgtacagcaa	caagttcaag	gcgcgggtga	tggtctcccg	caagaccccc	aatgggggtga	14340
cagtgcagaa	ggattatgat	ggtagtcagg	atgagctgaa	gtatgaatgg	gtggaatttg	14400
agctgcccga	aggcaacttc	tcggtgacca	tgaccatcga	cctgatgaac	aacgccatca	14460
tcgacaatta	cttggcgggtg	gggcggcaga	acgggggtgct	ggagagcgac	atcggcgtga	14520
agttcgacac	taggaacttc	aggctgggct	gggaccccgt	gaccgagctg	gtcatgcccc	14580
gggtgtacac	caacgaggct	ttccatcccc	atattgtctt	gctgcccggc	tcggggggtg	14640

acttcaccga	gagccgcctc	agcaacctgc	tgggcattcg	caagaggcag	cccttccagg	14700
aaggcttcca	gatcatgtac	gaggatctgg	aggggggcaa	catccccgcg	ctcctggatg	14760
tcgacgccta	tgagaaaagc	aaggaggatg	cagcagctga	agcaactgca	gccgtagcta	14820
ccgcctctac	cgaggtcagg	ggcgataatt	ttgcaagcgc	cgcagcagtg	gcagcggccg	14880
aggcggctga	aaccgaaagt	aagatagtca	ttcagccggg	ggagaaggat	agcaagaaca	14940
ggagctacaa	cgtactaccg	gacaagataa	acaccgccta	ccgcagctgg	tacctagcct	15000
acaactatgg	cgacccccgag	aagggcgtgc	gctcctggac	gctgctcacc	acctcggacg	15060
tcacctgcgg	cgtggagcaa	gtctactggg	cgctgcccga	catgatgcaa	gacccgggtca	15120
ccttccgctc	cacgcgtcaa	gttagcaact	acccgggtgg	gggcgccgag	ctcctgcccg	15180
tctactccaa	gagcttcttc	aacgagcagg	ccgtctactc	gcagcagctg	cgcgcttca	15240
cctcgcttac	gcacgtcttc	aaccgcttcc	ccgagaacca	gatcctcgtc	cgcccgcccg	15300
cgcccaccat	taccaccgtc	agtgaaaacg	ttcctgctct	cacagatcac	gggaccctgc	15360
cgctgcgcag	cagtatccgg	ggagtccagc	gcgtgaccgt	tactgacgcc	agacgccgca	15420
cctgccccta	cgtctacaag	gccctgggca	tagtcgcgcc	gcgcgtcctc	tcgagccgca	15480
ccttctaaat	gtccattctc	atctcgccca	gtaataacac	cggttggggc	ctgcgcgcgc	15540
ccagcaagat	gtacggaggc	gctcgccaac	gctccacgca	acaccccggtg	cgcggtgcgcg	15600
ggcacttccg	cgctccctgg	ggcgccctca	agggccgcgt	gcggtcgcgc	accaccgtcg	15660
acgacgtgat	cgaccaggtg	gtggccgacg	cgcgcaacta	cacccccgcc	gccgcgccccg	15720
tctccaccgt	ggacgccgtc	atcgacagcg	tggtagccga	cgcgcgccgg	tacgccccgcg	15780
ccaagagccg	gcggcggcgc	atcgcccggc	ggcaccggag	cacccccgcc	atgcgcgcgg	15840
cgcgagcctt	gctgcgcagg	gccaggcgca	cgggacgcag	ggccatgctc	agggcggcca	15900
gacgcgcggc	ttcaggcgcc	agcgccggca	ggacccggag	acgcgcggcc	acggcggcgg	15960
cagcggccat	cgccagcatg	tcccgcgcgc	ggcgagggaa	cgtgtactgg	gtgcgcgacg	16020
ccgccaccgg	tgtgcgcgtg	cccgtgcgca	ccgccccccc	tcgcacttga	agatgtttac	16080
ttcgcgatgt	tgatgtgtcc	cagcggcgag	gaggatgtcc	aagcgcaa	atcaaggaaga	16140
gatgctccag	gtcatcgcgc	ctgagatcta	cggccctgcg	gtggtgaagg	aggaaagaaa	16200
gccccgcaaa	atcaagcggg	tcaaaaagga	caaaaaggaa	gaagaaagtg	atgtggacgg	16260

attggtggag	tttgtgcgcg	agttcgcccc	ccggcggcgc	gtgcagtggc	gcgggcggaa	16320
ggtgcaaccg	gtgctgagac	ccggcaccac	cgtggtcttc	acgcccggcg	agcgtcccg	16380
caccgcttcc	aagcgctcct	acgacgaggt	gtacggggat	gatgatattc	tggagcaggc	16440
ggccgagcgc	ctgggcgagt	ttgcttacgg	caagcgcagc	cgttccgcac	cgaaggaaga	16500
ggcggtgtcc	atcccgtg	accacggcaa	ccccacgccg	agcctcaagc	ccgtgacctt	16560
gcagcaggtg	ctgccgaccg	cggcgccgcg	ccgggggttc	aagcgcgagg	gcgaggatct	16620
gtaccccacc	atgcagctga	tgggtgcccc	gcgccagaag	ctggaagacg	tgctggagac	16680
catgaaggtg	gacccggacg	tgagccccga	ggtcaaggtg	cggcccatca	agcaggtggc	16740
cccgggctg	ggcgtgcaga	ccgtggacat	caagattccc	acggagccca	tggaaacgca	16800
gaccgagccc	atgatcaagc	ccagcaccag	caccatggag	gtgcagacgg	atccctggat	16860
gccatcggct	cctagtcgaa	gaccccggcg	caagtacggc	gcggccagcc	tgctgatgcc	16920
caactacgcg	ctgcatcctt	ccatcatccc	cacgccgggc	taccgcggca	cgcgcttcta	16980
ccgcggtcat	accagcagcc	gccgccgcaa	gaccaccact	cgccgccgcc	gtcgccgcac	17040
cgccgctgca	accacccctg	ccgccctggt	gcggagagtg	taccgccgcg	gccgcgcacc	17100
tctgaccctg	ccgcgcgcgc	gctaccaccc	gagcatcgcc	atttaaactt	tcgcctgctt	17160
tgagatcaa	tggccctcac	atgccgcctt	cgcgttccca	ttacgggcta	ccgaggaaga	17220
aaaccgcgcc	gtagaaggct	ggcggggaac	gggatgcgtc	gccaccacca	ccggcggcgg	17280
cgcgccatca	gcaagcggtt	ggggggaggc	ttcctgccc	cgctgatccc	catcatcgcc	17340
gcggcgatcg	gggcgatccc	cggcattgct	tccgtggcgg	tgaggcctc	tcagcgccac	17400
tgagacacac	ttggaacat	cttgtaataa	accaatggac	tctgacgctc	ctggtcctgt	17460
gatgtgtttt	cgtagacaga	tggaagacat	caatttttcg	tccctggctc	cgcgacacgg	17520
cacgcggccg	ttcatgggca	cctggagcga	catcggcacc	agccaactga	acgggggcgc	17580
cttcaattgg	agcagtctct	ggagcgggct	taagaatttc	gggtccacgc	ttaaaacctt	17640
tggcagcaag	gcgtggaaca	gcaccacagg	gcaggcgctg	agggataagc	tgaaagagca	17700
gaacttccag	cagaaggtgg	tcgatgggct	cgcctcgggc	atcaacgggg	tgggtggacct	17760
ggccaaccag	gccgtgcagc	ggcagatcaa	cagccgcctg	gacccggtgc	cgcccgccgg	17820
ctccgtggag	atgccgcagg	tggaggagga	gctgcctccc	ctggacaagc	ggggcgagaa	17880

gcgacccccgc	cccgatgcgg	aggagacgct	gctgacgcac	acggacgagc	cgcccccgta	17940
cgaggaggcg	gtgaaactgg	gtctgcccac	cacgcggccc	atcgcgcccc	tggccaccgg	18000
ggtgctgaaa	ccgaaaagc	ccgcgaccct	ggacttgcct	cctccccagc	cttcccggcc	18060
ctctacagtg	gctaagcccc	tgccgccggt	ggccgtggcc	cgcgcgcgac	ccggggggcac	18120
cgcccgccct	catgcgaact	ggcagagcac	tctgaacagc	atcgtgggtc	tgggagtgca	18180
gagtgtgaag	cgccgccgct	gctattaaac	ctaccgtagc	gcttaacttg	cttgtctgtg	18240
tgtgtatgta	ttatgtcgcc	gccgccgctg	tccaccagaa	ggaggagtga	agaggcgcgt	18300
cgccgagttg	caagatggcc	accccatcga	tgctgcccc	gtgggcgtac	atgcacatcg	18360
ccggacagga	cgcttcggag	tacctgagtc	cgggtctggt	gcagtttgcc	cgcgccacag	18420
acacctactt	cagtctgggg	aacaagttta	ggaacccac	ggtggcgccc	acgcacgatg	18480
tgaccaccga	ccgcagccag	cggctgacgc	tgcgcttcgt	gcccgtggac	cgcgaggaca	18540
acacctactc	gtacaaagtg	cgctacacgc	tggccgtggg	cgacaaccgc	gtgctggaca	18600
tggccagcac	ctactttgac	atccgcggcg	tgctggatcg	gggccctagc	ttcaaaccct	18660
actccggcac	cgcttacaac	agtcctggccc	ccaagggagc	acccaacact	tgtcagtgga	18720
catataaagc	cgatggtgaa	actgccacag	aaaaaaccta	tacatatgga	aatgcacccg	18780
tgcagggcat	taacatcaca	aaagatggta	ttcaacttgg	aactgacacc	gatgatcagc	18840
caatctacgc	agataaaacc	tatcagcctg	aacctcaagt	gggtgatgct	gaatggcatg	18900
acatcactgg	tactgatgaa	aagtatggag	gcagagctct	taagcctgat	acaaaaatga	18960
agccttgтта	tggttctttt	gccaagccta	ctaataaaga	aggaggtcag	gcaaatgtga	19020
aaacaggaac	aggcactact	aaagaatatg	acatagacat	ggctttcttt	gacaacagaa	19080
gtgcggctgc	tgctggccta	gctccagaaa	ttgttttgta	tactgaaaat	gtggatttgg	19140
aaactccaga	tacccatatt	gtatacaaag	caggcacaga	tgacagcagc	tcttctatta	19200
atttgggtca	gcaagccatg	cccaacagac	ctaactacat	tggtttcaga	gacaacttta	19260
tcgggctcat	gtactacaac	agcactggca	atatgggggt	gctggccggt	caggcttctc	19320
agctgaatgc	tgtggttgac	ttgcaagaca	gaaacaccga	gctgtcctac	cagctcttgc	19380
ttgactctct	gggtgacaga	acccggtatt	tcagtatgtg	gaatcaggcg	gtggacagct	19440
atgatcctga	tgtgcgcat	attgaaaatc	atggtgtgga	ggatgaactt	cccaactatt	19500

gtttccctct	ggatgctgtt	ggcagaacag	atacttatca	gggaattaag	gctaattgaa	19560
ctgatcaaac	cacatggacc	aaagatgaca	gtgtcaatga	tgctaattgag	ataggcaagg	19620
gtaatccatt	cgccatggaa	atcaacatcc	aagccaacct	gtggagggaac	ttcctctacg	19680
ccaacgtggc	cctgtacctg	cccgactctt	acaagtacac	gccggccaat	gttaccctgc	19740
ccaccaacac	caacacctac	gattacatga	acggccgggt	ggtggcgccc	tcgctggtgg	19800
actcctacat	caacatcggg	gcgcgctggt	cgctggatcc	catggacaac	gtgaaccctt	19860
tcaaccacca	ccgcaatgcg	gggctgcgct	accgctccat	gctcctgggc	aacgggcgct	19920
acgtgccctt	ccacatccag	gtgccccaga	aatttttcgc	catcaagagc	ctcctgctcc	19980
tgcccgggtc	ctacacctac	gagtgggaact	tccgcaagga	cgtcaacatg	atcctgcaga	20040
gtccctcgg	caacgacctg	cgcacggacg	gggcctccat	ctccttcacc	agcatcaacc	20100
tctacgccac	cttcttcccc	atggcgcaca	acacggcctc	cacgctcgag	gcatgctgc	20160
gcaacgacac	caacgaccag	tccttcaacg	actacctctc	ggcggccaac	atgctctacc	20220
ccatcccggc	caacgccacc	aacgtgcca	tctccatccc	ctcgcgcaac	tgggccgcct	20280
tccgcgggtg	gtccttcacg	cgtctcaaga	ccaaggagac	gccctcgctg	ggctccgggt	20340
tcgacccta	cttcgtctac	tcgggctcca	tcccctacct	cgacggcacc	ttctacctca	20400
accacacctt	caagaaggct	tccatcacct	tcgactcctc	cgtcagctgg	cccggcaacg	20460
accggctcct	gacgccaac	gagttcgaaa	tcaagcgcac	cgtcgacggc	gagggctaca	20520
acgtggcca	gtgcaacatg	accaaggact	ggttcctggt	ccagatgctg	gcccactaca	20580
acatcggcta	ccagggttc	tacgtgcccg	agggtacaa	ggaccgcatg	tactccttct	20640
tccgcaactt	ccagcccatg	agccgccagg	tggtggacga	ggtcaactac	aaggactacc	20700
aggccgtcac	cctggcctac	cagcacaaca	actcgggctt	cgtcggctac	ctcgcgcca	20760
ccatgcgcca	gggccagccc	taccccgcca	actacccta	cccgtcatc	ggcaagagcg	20820
ccgtcaccag	cgtcaccag	aaaaagttcc	tctgcgacag	ggtcatgtgg	cgcattcccct	20880
tctccagcaa	cttcattgtc	atgggcgcgc	tcaccgacct	cggccagaac	atgctctatg	20940
ccaactccgc	ccacgcgcta	gacatgaatt	tcgaagtcga	ccccatggat	gagtccaccc	21000
ttctctatgt	tgtcttcgaa	gtcttcgacg	tcgtccgagt	gcaccagccc	caccgcggcg	21060
tcatcgaggc	cgtctacctg	cgcaccccct	tctcggccgg	taacgccacc	acctaagctc	21120

ttgcttcttg caagccatgg ccgctgggctc cggcgagcag gagctcaggg ccatcatccg	21180
cgacctgggc tgcgggacct acttcctggg caccttcgat aagcgcttcc cgggattcat	21240
ggccccgcac aagctggcct gcgccatcgt caacacggcc ggccgcgaga ccgggggga	21300
gcactggctg gccttcgcct ggaacccgcg ctcgaacacc tgctacctct tcgacctt	21360
cgggttctcg gacgagcgcc tcaagcagat ctaccagttc gagtacgagg gcctgctgcg	21420
ccgcagcgcc ctggccaccg aggaccgctg cgtcaccttg gaaaagtcca ccagaccgt	21480
gcagggtccg cgctcggccg cctgcgggct cttctgctgc atgttcctgc acgccttcgt	21540
gcactggccc gaccgcccc aaggacaaga cccaccatg aacttgctga cgggggtgcc	21600
caacggcatg ctccagtcgc ccaggtgga acccaccctg cgccgcaacc aggaggcgct	21660
ctaccgcttc ctcaactccc actccgccta ctttcgctcc caccgcgcgc gcatcgagaa	21720
ggccaccgcc ttcgaccgca tgaatcaaga catgtaaacc gtgtgtgtat gttaaatgtc	21780
tttaataaac agcactttca tgttacacat gcatctgaga tgatttatat agaaatcgaa	21840
agggttctgc cgggtctcgg catggcccgc gggcaggagc acgttgcgga actggtactt	21900
ggccagccac ttgaactcgg ggatcagcag tttgggcagc ggggtgtcgg ggaaggagtc	21960
ggtccacagc ttccgcgtca gttgcagggc gccagcagg tcgggcgcgg agatcttgaa	22020
atcgagttg ggacccgcgt tctgcgcgcg ggagttgcgg tacacggggt tgcagcactg	22080
gaacaccatc agggccgggt gcttcacgct cgccagcacc gtcgcgtcgg tgatgctctc	22140
cacgtcagg tcctcggcgt tggccatccc gaaggggggc atcttgagg tctgccttcc	22200
catggtgggc acgcaccgg gcttgtggtt gcaatcgcag tgcaggggga tcagcatcat	22260
ctgggcctgg tcggcgttca tccccgggta catggccttc atgaaagcct ccaattgcct	22320
gaacgcctgc tgggccttgg ctccctcggg gaagaagacc ccgcaggact tgctagagaa	22380
ctggttggtg gcgcaccgg cgtcgtgcac gcagcagcg gcgtcgttgt tggccagctg	22440
caccacgctg cgcggccagc ggttctgggt gatcttggcc cggtcggggt tctccttcag	22500
cgcgcgctgc ccgttctcgc tcgccacatc catctcgatc atgtgctcct tctggatcat	22560
ggtggtcccg tgcaggcacc gcagcttgcc ctcggcctcg gtgcacccgt gcagccacag	22620
cgcgcacccg gtgcactccc agttcttgtg ggcgatctgg gaatgcgcgt gcacgaagcc	22680
ctgcaggaag cggcccatca tgggtggtcag ggtcttggtg ctagtgaagg tcagcggaat	22740

gccgcggtgc	tcctcgttga	tgtacaggtg	gcagatgcgg	cggtacacct	cgccctgctc	22800
gggcatcagc	tggaagttgg	ctttcaggtc	ggtctccacg	cggtagcggg	ccatcagcat	22860
agtcattgatt	tccataacct	tctcccaggc	cgagacgatg	ggcaggctca	tagggttctt	22920
caccatcatc	ttagcgctag	cagccgcggc	caggggggtc	ctctcgtcca	gggtctcaaa	22980
gctccgcttg	ccgtccttct	cggtgatccg	caccgggggg	tagctgaagc	ccacggccgc	23040
cagctcctcc	tcggcctgtc	tttcgtcctc	gctgtcctgg	ctgacgtcct	gcaggaccac	23100
atgcttggtc	ttgcgggggt	tcttcttggg	cggcagcggc	ggcggagatg	ttggagatgg	23160
cgagggggag	cgcgagttct	cgctcaccac	tactatctct	tcctcttctt	ggtccgaggc	23220
cacgcggcgg	taggtatgtc	tcttcggggg	cagaggcgga	ggcgacgggc	tctcgccgcc	23280
gcgacttggc	ggatggctgg	cagagcccct	tccgcgttcg	gggggtgcgt	cccggcggcg	23340
ctctgactga	cttcctccgc	ggccggccat	tgtgttctcc	tagggaggaa	caacaagcat	23400
ggagactcag	ccatcgccaa	cctcgccatc	tgccccacc	gccgacgaga	agcagcagca	23460
gcagaatgaa	agcttaaccg	ccccgccgcc	cagccccgcc	acctccgacg	cggccgtccc	23520
agacatgcaa	gagatggagg	aatccatcga	gattgacctg	ggctatgtga	cgccgcgga	23580
gcacgaggag	gagctggcag	tgcgcttttc	acaagaagag	atacaccaag	aacagccaga	23640
gcaggaagca	gagaatgagc	agagtcaggc	tgggctcgag	catgacggcg	actacctcca	23700
cctgagcggg	ggggaggacg	cgctcatcaa	gcatctggcc	cggcaggcca	ccatcgtcaa	23760
ggatgcgctg	ctcgaccgca	ccgaggtgcc	cctcagcgtg	gaggagctca	gccgcgccta	23820
cgagttgaac	ctcttctcgc	cgcgcttgcc	ccccaaagcg	cagcccaatg	gcacctgcga	23880
gcccacccg	cgcttcaact	tctacccggt	cttcgcgggt	cccaggccc	tggccaccta	23940
ccacatcttt	ttcaagaacc	aaaagatccc	cgtctcctgc	cgcgccaacc	gcacccgcgc	24000
cgacgccctt	ttcaacctgg	gtcccggcgc	ccgcctacct	gatatcgctt	ccttggaaga	24060
ggttcccaag	atcttcgagg	gtctgggcag	cgacgagact	cgggccgcga	acgctctgca	24120
aggagaagga	ggagagcatg	agcaccacag	cgccctgggt	gagttggaag	gcgacaacgc	24180
gcggctggcg	gtgctcaaac	gcacggtcga	gctgacccat	ttcgcttacc	cggctctgaa	24240
cctgcccccc	aaagtcatga	gcgcggtcat	ggaccaggtg	ctcatcaagc	gcgcgtcgcc	24300
catctccgag	gacgagggca	tgcaagactc	cgaggagggc	aagcccgtgg	tcagcgacga	24360

gcagctggcc	cggtggctgg	gtcctaattgc	tagtccccag	agtttggaag	agcggcgcaa	24420
actcatgatg	gccgtggctc	tggtgaccgt	ggagctggag	tgcttgcgcc	gcttcttcgc	24480
cgacgcggag	accctgcgca	aggtcgagga	gaacctgcac	tacctcttca	ggcacgggtt	24540
cgtgcgccag	gcctgcaaga	tctccaacgt	ggagctgacc	aacctgggtct	cctacatggg	24600
catcttgcac	gagaaccgcc	tggggcagaa	cgtgctgcac	accaccctgc	gcggggaggc	24660
ccggcgcgac	tacatccgcg	actgcgtcta	cctctacctc	tgccacacct	ggcagacggg	24720
catgggcgtg	tggcagcagt	gtctggagga	gcagaacctg	aaagagctct	gcaagctcct	24780
gcagaagaac	ctcaagggtc	tgtggaccgg	gttcgacgag	cgcaccaccg	cctcggacct	24840
ggccgacctc	atthttccccg	agcgcctcag	gctgacgctg	cgcaacggcc	tgcccgactt	24900
tatgagccaa	agcatgttgc	aaaactttcg	ctctttcatc	ctcgaacgct	ccggaatcct	24960
gcccgccacc	tgctccgcgc	tgccctcgga	cttcgtgccg	ctgaccttcc	gcgagtgcgc	25020
cccgccgctg	tggagccact	gctacctgct	gcgcctggcc	aactacctgg	cctaccactc	25080
ggacgtgatc	gaggacgtca	gcggcgaggg	cctgctcgag	tgccactgcc	gctgcaacct	25140
ctgcacgccg	caccgctccc	tggcctgcaa	ccccagctg	ctgagcgaga	cccagatcat	25200
cggcaccttc	gagttgcaag	ggcccagcga	aggcgagggt	tcagccgcca	aggggggtct	25260
gaaactcacc	ccggggctgt	ggacctcggc	ctacttgccg	aagttcgtgc	ccgaggacta	25320
ccatcccttc	gagatcaggt	tctacgagga	ccaatcccat	ccgcccagg	ccgagctgtc	25380
ggcctgcgtc	atcaccagc	gggcgatcct	ggcccaattg	caagccatcc	agaaatcccg	25440
ccaagaattc	ttgctgaaaa	agggccgcgg	ggtctacctc	gacccccaga	ccggtgagga	25500
gctcaacccc	ggcttcccc	aggatgcccc	gaggaaacaa	gaagctgaaa	gtggagctgc	25560
cgcccgtgga	ggatttggag	gaagactggg	agaacagcag	tcaggcagag	gaggaggaga	25620
tgagggaaga	ctgggacagc	actcaggcag	aggaggacag	cctgcaagac	agtctggagg	25680
aagacgagga	ggaggcagag	gaggaggtgg	aagaagcagc	cgccgccaga	ccgtcgtcct	25740
cggcggggga	gaaagcaagc	agcacggata	ccatctccgc	tccgggtcgg	ggtcccgtc	25800
gaccacacag	tagatgggac	gagaccggac	gattcccgaa	ccccaccacc	cagaccggta	25860
agaaggagcg	gcagggatac	aagtcctggc	gggggcacaa	aaacgccatc	gtctcctgct	25920
tgagggcctg	cgggggcaac	atctccttca	cccggcgcta	cctgctcttc	caccgcgggg	25980

tgaactttcc	ccgcaacatc	ttgcattact	accgtcacct	ccacagcccc	tactacttcc	26040
aagaagaggc	agcagcagca	gaaaaagacc	agcagaaaaac	cagcagctag	aaaatccaca	26100
gcggcggcag	caggtggact	gaggatcgcg	gcgaacgagc	cggcgcaaac	ccgggagctg	26160
aggaaccgga	tctttccac	cctctatgcc	atcttccagc	agagtcgggg	gcaggagcag	26220
gaactgaaag	tcaagaaccg	ttctctgcgc	tcgctcacc	gcagttgtct	gtatcacaag	26280
agcgaagacc	aacttcagcg	cactctcgag	gacgccgagg	ctctcttcaa	caagtactgc	26340
gcgctcactc	ttaaagagta	gcccgcgccc	gcccgatcg	agaaaaaggc	gggaattacg	26400
tcacctgtgc	ccttcgccct	agccgcctcc	acccatcatc	atgagcaaag	agattcccac	26460
gccttacatg	tggagctacc	agccccagat	gggcctggcc	gccggtgccg	cccaggacta	26520
ctccaccgc	atgaattggc	tcagcgccgg	gcccgcgatg	atctcacggg	tgaatgacat	26580
ccgcgcccac	cgaaccaga	tactcctaga	acagtcagcg	ctcaccgcca	cgccccgcaa	26640
tcacctcaat	ccgcgtaatt	ggcccgccgc	cctgggtgtac	caggaaattc	cccagcccac	26700
gaccgtacta	cttccgcgag	acgcccaggc	cgaagtccag	ctgactaact	caggtgtcca	26760
gctggcgggc	ggcgccaccc	tgtgtcgtca	ccgccccgct	cagggtataa	agcggctggt	26820
gatccggggc	agaggcacac	agctcaacga	cgagggtggtg	agctcttcgc	tgggtctgcg	26880
acctgacgga	gtcttccaac	tcgccggatc	ggggagatct	tccttcacgc	ctcgtcaggc	26940
cgtcctgact	ttggagagtt	cgtcctcgca	gccccgctcg	ggtggcatcg	gcactctcca	27000
gttcgtggag	gagttcactc	cctcggtcta	cttcaacccc	ttctccggct	ccccggcca	27060
ctaccggac	gagttcatcc	cgaacttcga	cgccatcagc	gagtcggtgg	acggctacga	27120
ttgaatgtcc	catggtggcg	cagctgacct	agctcggctt	cgacacctgg	accactgccg	27180
ccgcttccgc	tgcttcgctc	gggatctcgc	cgagtttgcc	tactttgagc	tgcccgagga	27240
gcaccctcag	ggcccggccc	acggagtgcg	gatcgtcgtc	gaagggggcc	tcgactcca	27300
cctgcttcgg	atcttcagcc	agcgtccgat	cctggtcgag	cgcgagcaag	gacagaccct	27360
tctgactctg	tactgcatct	gcaaccaccc	cggcctgcat	gaaagtcttt	gttgtctgct	27420
gtgtactgag	tataataaaa	gctgagatca	gcgactactc	cggacttccg	tgtgttcctg	27480
aatccatcaa	ccagtctttg	ttcttcaccg	ggaacgagac	cgagctccag	ctccagtgtg	27540
agccccacaa	gaagtacctc	acctggctgt	tccagggtc	cccgatcgcc	gttgtcaacc	27600

actgcgacaa cgacggagtc ctgctgagcg gccctgccaa cttactttt tccacccgca	27660
gaagcaagct ccagctcttc caacccttcc tccccgggac ctatcagtgc gtctcgggac	27720
cctgccatca caccttcac ctgatcccga ataccacagc gtcgctcccc gctactaaca	27780
acaaaactaa cctccaccaa cgccaccgtc gcgacctttc tgaatctaata actaccaccc	27840
acaccggagg tgagctccga ggtcaaccaa cctctgggat ttactacggc ccctgggagg	27900
tggttgggtt aatagcgcta ggcctagttg cgggtgggct tttggttctc tgctacctat	27960
acctcccttg ctgttcgtac ttagtggtgc tgtgttgctg gtttaagaaa tggggaagat	28020
caccctagtg agctgcggtg cgctgggtggc ggtgttgctt tcgattgtgg gactgggcgg	28080
tgcggctgta gtgaaggaga aggccgatcc ctgcttgcat ttcaatccca acaaatgcca	28140
gctgagtttt cagcccgatg gcaatcggtg cgcggtactg atcaagtgcg gatgggaatg	28200
cgagaacgtg agaatcgagt acaataacaa gactcggaac aatactctcg cgtccgtgtg	28260
gcagcccggg gaccccgagt ggtacaccgt ctctgtcccc ggtgctgacg gctccccgcg	28320
caccgtgaat aatactttca tttttgcgca catgtgcgac acggtcatgt ggatgagcaa	28380
gcagtacgat atgtggcccc ccacgaagga gaacatcgtg gtcttctcca tcgcttacag	28440
cctgtgcacg gcgctaata caagctatcgt gtgcctgagc attcacatgc tcatcgctat	28500
tcgccccaga aataatgccg aaaaagaaaa acagccataa cgtttttttt cacacctttt	28560
tcagaccatg gcctctgtta aatttttgct tttatttgcc agtctcattg ccgtcattca	28620
tggaatgagt aatgagaaaa ttactattta cactggcact aatcacacat tgaaaggtcc	28680
agaaaaagcc acagaagttt catggtattg ttattttaat gaatcagatg tatctactga	28740
actctgtgga aacaataaca aaaaaaatga gagcattact ctcacatcaagt ttcaatgtgg	28800
atctgactta accctaatta acatcactag agactatgta ggtatgtatt atggaactac	28860
agcaggcatt tcggacatgg aattttatca agtttctgtg tctgaaccca ccacgcctag	28920
aatgaccaca accacaaaaa ctacacctgt taccactatg cagctcacta ccaataacat	28980
ttttgcatg cgtcaaatgg tcaacaatag cactcaaccc accccaccca gtgaggaaat	29040
tcccaaatcc atgattggca ttattgttgc tgtagtggtg tgcatgttga tcatgcctt	29100
gtgcatggtg tactatgcct tctgctacag aaagcacaga ctgaacgaca agctggaaca	29160
cttactaagt gttgaatttt aattttttag aaccatgaag atcctaggcc ttttaatttt	29220

ttctatcatt	acctctgctc	tatgcaattc	tgacaatgag	gacgttactg	tcgttgtcgg	29280
atcaaattat	acactgaaag	gtccagcgaa	gggtatgctt	tcgtgggtatt	gctattttgg	29340
atctgacact	acagaaactg	aattatgcaa	tcttaagaat	ggcaaaattc	aaaattctaa	29400
aattaacaat	tatatatgca	atggtactga	tctgatactc	ctcaatatca	cgaaatcata	29460
tgctggcagt	tacacctgcc	ctggagatga	tgctgacagt	atgatttttt	acaaagtaac	29520
tgttgttgat	cccactactc	cacctccacc	caccacaact	actcacacca	cacacacaga	29580
tcaaaccgca	gcagaggagg	cagcaaagtt	agccttgtag	gtccaagaca	gttcatttgt	29640
tggcattacc	cctacacctg	atcagcggtg	tccggggctg	ctagtcagcg	gcattgtcgg	29700
tgtgctttcg	ggattagcag	tcataatcat	ctgcatgttc	atttttgctt	gctgctatag	29760
aaggctttac	cgacaaaaat	cagaccact	gctgaacctc	tatgtttaat	tttttcaga	29820
gtcatgaagg	cagttagcgc	tctagttttt	tgttctttga	ttggcattgt	tttttgcaat	29880
cctattccta	aagttagctt	tattaaagat	gtgaatgtta	ctgagggggg	caatgtgaca	29940
ctggtaggtg	tagagggtgc	tgaaaacacc	acctggacaa	aataccacct	caatgggtgg	30000
aaagatattt	gcaattggag	tgtattagtt	tatacatgtg	aggagagtaa	tcttaccatt	30060
gtcaatgcca	cctcagctca	aatggtaga	attcaaggac	aaagtgtcag	tgtatcta	30120
gggtatttta	cccaacatac	ttttatctat	gacgttaaag	tcataccact	gcctacgcct	30180
agcccaccta	gcactaccac	acagacaacc	cacactacac	agacaaccac	atacagtaca	30240
ttaaatacgc	ctaccaccac	tacagcagca	gaggttgcca	gctcgtctgg	ggtccgagtg	30300
gcatttttga	tgttggcccc	atctagcagt	cccactgcta	gtaccaatga	gcagactact	30360
gaatttttgt	ccactgtcga	gagccacacc	acagctacct	ccagtgcctt	ctctagcacc	30420
gccaatctct	cctcgctttc	ctctacacca	atcagtcccg	ctactactcc	tagccccgct	30480
cctcttccca	ctcccctgaa	gcaaacagac	ggcggcatgc	aatggcagat	caccctgctc	30540
attgtgatcg	ggttgggtcat	cctggccgtg	ttgctctact	acatcttctg	ccgccgcatt	30600
cccaacgcgc	accgcaagcc	ggtctacaag	cccatcattg	tcgggcagcc	ggagccgctt	30660
caggtggaag	ggggtctaag	gaatcttctc	ttctctttta	cagtatgggtg	attgaactat	30720
gattcctaga	caattcttga	tcactattct	tatctgcctc	ctccaagtct	gtgccaccct	30780
cgctctggtg	gccaacgcca	gtccagactg	tattggggcc	ttcgctcct	acgtgctctt	30840

tgcccttcacc	acctgcatct	gctgctgtag	catagtctgc	ctgcttatca	ccttcttcca	30900
gttcattgac	tggatctttg	tgcgcatcgc	ctacctgcgc	caccaccccc	agtaccgcga	30960
ccagcgagtg	gcgcggctgc	tcaggctcct	ctgataagca	tgcgggctct	gctacttctc	31020
gcgcttctgc	tgttagtgtc	ccccgtccc	gtcgaccccc	ggccccccac	ccagtcccc	31080
gaggaggtcc	gcaaatgcaa	attccaagaa	ccctggaaat	tcctcaaata	ctaccgccaa	31140
aaatcagaca	tgcatcccag	ctggatcatg	atcattggga	tcgtgaacat	tctggcctgc	31200
accctcatct	cctttgtgat	ttaccctgc	tttgactttg	gttggaactc	gccagaggcg	31260
ctctatctcc	cgctgaacc	tgacacacca	ccacagcaac	ctcaggcaca	cgactacca	31320
ccactacagc	ctaggccaca	atacatgccc	atattagact	atgaggccga	gccacagcga	31380
cccatgctcc	ccgctattag	ttacttcaat	ctaaccggcg	gagatgactg	accactggc	31440
caacaacaac	gtcaacgacc	ttctcctgga	catggacggc	cgcgctcgg	agcagcgact	31500
cgcccaactt	cgcatcgc	agcagcagga	gagagccgtc	aaggagctgc	aggatgcggt	31560
ggccatccac	cagtgaaga	gaggcatctt	ctgcctgggtg	aaacaggcca	agatctccta	31620
cgaggtcact	ccaaacgacc	atcgctctc	ctacgagctc	ctgcagcagc	gccagaagtt	31680
cacctgcctg	gtcggagtca	accccatcgt	catcaccag	cagtctggcg	ataccaaggg	31740
gtgcatccac	tgctcctgcg	actccccga	ctgcgtccac	actctgatca	agaccctctg	31800
cggcctccgc	gacctcctcc	ccatgaacta	atcacccct	tatccagtga	aataaagatc	31860
atattgatga	tgattttaca	gaaataaaaa	ataatcattt	gatttgaaat	aaagatacaa	31920
tcatattgat	gatttgagtt	taacaaaaaa	ataaagaatc	acttacttga	aatctgatac	31980
caggtctctg	tccatgtttt	ctgccaacac	cacttcactc	ccctcttccc	agctctggta	32040
ctgcaggccc	cggcgggctg	caaacttcct	ccacacgctg	aaggggatgt	caaattcctc	32100
ctgtccctca	atcttcattt	tatcttctat	cagatgtcca	aaaagcgcgt	ccgggtggat	32160
gatgacttcg	acccgtcta	cccctacgat	gcagacaacg	caccgaccgt	gcccttcatc	32220
aaacccccct	tcgtctcttc	agatggattc	caagagaagc	ccctgggggt	gttgtccctg	32280
cgactggccg	acccgtcac	caccaagaac	ggggaaatca	ccctcaagct	gggagagggg	32340
gtggacctcg	attcctcggg	aaaactcatc	tccaacacgg	ccaccaaggc	cgccgcccct	32400
ctcagttttt	ccaacaacac	catttcctt	aacatggatc	acccctttta	cactaaagat	32460

ggaaaattat ccttacaagt ttctccacca ttaaataatac tgagaacaag cattctaaac	32520
acactagctt taggttttgg atcagggttta ggactccgtg gctctgcctt ggacgtacag	32580
ttagtctctc cacttacatt tgatactgat ggaaacataa agcttacctt agacagaggt	32640
ttgcatgtta caacaggaga tgcaattgaa agcaacataa gctgggctaa aggtttaaaa	32700
tttgaagatg gagccatagc aaccaacatt ggaaatgggt tagagtttgg aagcagtagt	32760
acagaaacag gtgttgatga tgcttaccca atccaagtta aacttggatc tggccttagc	32820
tttgacagta caggagccat aatggctggt aacaaagaag acgataaact cactttgtgg	32880
acaacacctg atccatcacc aaactgtcaa atactcgcag aaaatgatgc aaaactaaca	32940
ctttgcttga ctaaattgtgg tagtcaaata ctggccactg tgtcagtctt agttgtagga	33000
agtggaaacc taaaccccat tactggcacc gtaagcagtg ctcaggtgtt tctacgtttt	33060
gatgcaaacg gtgttctttt aacagaacat tctacactaa aaaaatactg ggggtatagg	33120
caggagata gcatagatgg cactccatat accaatgctg taggattcat gccaattta	33180
aaagcttatc caaagtcaca aagttctact actaaaaata atatagtagg gcaagtatac	33240
atgaatggag atgtttcaaa acctatgctt ctactataa ccctcaatgg tactgatgac	33300
agcaacagta catattcaat gtcattttca tacacctgga ctaatggaag ctatgttgga	33360
gcaacatttg gggctaactc ttataccttc tcatacatcg cccaagaatg aacactgtat	33420
cccaccctgc atgccaaacc ttcccacccc actctgtgga acaaactctg aaacacaaaa	33480
taaaataaag ttcaagtgtt ttattgattc aacagtttta caggattcga gcagttattt	33540
ttcctccacc ctcccaggac atggaataca ccaccctctc ccccgacaca gccttgaaca	33600
tctgaatgcc attggtgatg gacatgcttt tggctctcac gttccacaca gtttcagagc	33660
gagccagtct cgggtcggtc agggagatga aaccctccgg gcactccgc atctgcacct	33720
cacagctcaa cagctgagga ttgtcctcgg tggtcgggat cacggttatc tggaagaagc	33780
agaagagcgg cggtgggaat catagtccgc gaacgggatc ggccggtggt gtcgcatcag	33840
gccccgcagc agtcgctgcc gccgccgctc cgtcaagctg ctgctcaggg ggtccgggtc	33900
cagggactcc ctcagcatga tgcccacggc cctcagcatc agtcgtctgg tgcggcgggc	33960
gcagcagcgc atgcggatct cgctcaggtc gctgcagtac gtgcaacaca gaaccaccag	34020
gttgttcaac agtccatagt tcaacacgct ccagccgaaa ctcacgcgg gaaggatgct	34080

accacacgtgg	ccgtcgtacc	agatcctcag	gtaaatcaag	tggtgcccc	tccagaacac	34140
gctgcccacg	tacatgatct	ccttgggcat	gtggcggttc	accacctccc	ggtaccacat	34200
caccctctgg	ttgaacatgc	agccccggat	gatcctgcgg	aaccacaggg	ccagcaccgc	34260
ccgccccgcc	atgcagcgaa	gagaccccgg	gtcccggcaa	tggcaatgga	ggacccaccg	34320
ctcgtacccg	tggatcatct	gggagctgaa	caagtctatg	ttggcacagc	acaggcatat	34380
gctcatgcat	ctcttcagca	ctctcaactc	ctcgggggtc	aaaaccatat	cccagggcac	34440
ggggaactct	tgcaggacag	cgaaccccg	agaacagggc	aatcctcgca	cagaacttac	34500
attgtgcatg	gacagggtat	cgcaatcagg	cagcaccggg	tgatcctcca	ccagagaagc	34560
gcgggtctcg	gtctcctcac	agcgtggtaa	gggggcccgc	cgatacgggt	gatggcggga	34620
cgcggctgat	cgtgttcg	accgtgtcat	gatgcagttg	ctttcggaca	ttttcgtact	34680
tgctgtagca	gaacctggtc	cgggcgctgc	acaccgatcg	ccggcggcgg	tctcggcgct	34740
tggaacgctc	ggtgttgaaa	ttgtaaaaca	gccactctct	cagaccgtgc	agcagatcta	34800
gggcctcagg	agtgatgaag	atcccatcat	gcctgatggc	tctgatcaca	tcgaccaccg	34860
tggaatgggc	cagaccagc	cagatgatgc	aattttgttg	ggtttcgggtg	acggcggggg	34920
agggaagaac	aggagaacc	atgattaact	tttaatccaa	acggtctcgg	agtacttcaa	34980
aatgaagatc	gcggagatgg	cacctctcgc	ccccgctgtg	ttggtggaaa	ataacagcca	35040
ggtcaaagg	gatacggttc	tcgagatggt	ccacgggtggc	ttccagcaaa	gcctccacgc	35100
gcacatccag	aaacaagaca	atagcgaaag	cgggagggtt	ctctaattcc	tcaatcatca	35160
tgttacactc	ctgcaccatc	cccagataat	tttcattttt	ccagccttga	atgattcgaa	35220
ctagttcctg	aggtaaatac	aagccagcca	tgataaagag	ctcgcgcaga	gcgccctcca	35280
ccggcattct	taagcacacc	ctcataattc	caagatatct	tgctcctggt	tcacctgcag	35340
cagattgaca	agcggaatat	caaaatctct	gccgcgatcc	ctgagctcct	ccctcagcaa	35400
taactgtaag	tactctttca	tatcctctcc	gaaattttta	gccataggac	caccaggaat	35460
aagattaggg	caagccacag	tacagataaa	ccgaagtcct	ccccagttag	cattgcaaaa	35520
tgcaagactg	ctataagcat	gctggctaga	cccggtgata	tcttccagat	aactggacag	35580
aaaatcgccc	aggcaatttt	taagaaaatc	aacaaaagaa	aaatcctcca	ggtggacgtt	35640
tagagcctcg	ggaacaacga	tgaagtaa	gcaagcgggtg	cgttccagca	tggttagtta	35700

gctgatctgt agaaaaaaca aaaatgaaca ttaaaccatg ctagcctggc gaacaggtgg	35760
gtaaatcggt ctctccagca ccaggcaggc cacgggggtct ccggcgcgac cctcgtaaaa	35820
attgtcgcta tgattgaaaa ccatcacaga gagacgttcc cgggtggccgg cgtgaatgat	35880
tcgacaagat gaatacaccc ccggaacatt ggcgtccgcg agtgaaaaaa agcgcccgag	35940
gaagcaataa ggcactacaa tgctcagtct caagtccagc aaagcgatgc catgcggatg	36000
aagcacaaaa ttctcaggtg cgtacaaaat gtaattactc ccctcctgca caggcagcaa	36060
agccccgat ccctccaggt acacatacaa agcctcagcg tccatagctt accgagcagc	36120
agcacacaac aggcgcaaga gtcagagaaa ggctgagctc taacctgtcc acccgctctc	36180
tgctcaatat atagcccaga tctacactga cgtaaaggcc aaagtctaaa aatacccgcc	36240
aaataatcac acacgcccag cacacgcca gaaaccggtg acacactcaa aaaaatacgc	36300
gcacttcctc aaacgcccac aactgccgtc atttccgggt tcccacgcta cgtcatcaaa	36360
acacgacttt caaattccgt cgaccgttaa aaacgtcacc cgccccgcc ctaacggtcg	36420
cccgctcttc agccaatcag cgccccgcat ccccaaattc aaacacctca tttgcatatt	36480
aacgcgcaca aaaagtttga ggtatattat tgatgatgg	36519

<210> 11

<211> 31867

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 11

ccatcttcaa taatatacct caaacttttt gtgcgcgtta atatgcaaat gaggcgtttg	60
aatttgggga ggaagggcgg tgattggctg agggatgagc gaccgttagg ggcggggcga	120
gtgacgtttt gatgacgtgg ttgcgaggag gagccagttt gcaagttctc gtgggaaaag	180
tgacgtcaaa cgaggtgtgg tttgaacacg gaaataactca attttcccgc gctctctgac	240
aggaaatgag gtgtttctgg gcggatgcaa gtgaaaacgg gccattttcg cgcgaaaact	300
gaatgaggaa gtgaaaatct gagtaatttc gcgtttatgg caggaggag tatttgccga	360
gggccgagta gactttgacc gattacgtgg gggtttcgat taccgtgttt ttcacctaaa	420

tttccgcgta cggtgtcaaa gtccggtggt tttacgtagg tgtcagctga tcgccagggt	480
atttaaacct gcgctctcca gtcaagaggc cactcttgag tgccagcgag aagagttttc	540
tcctccgcgc cgcgagtcag atctacactt tgaaagtagg gataacaggg taatgacatt	600
gattattgac tagttgttaa tagtaatcaa ttacgggggtc attagttcat agcccatata	660
tggagttccg cgttacataa cttacggtaa atggcccgcc tggctgaccg cccaacgacc	720
cccgcccatt gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc	780
attgacgtca atgggtggag tatttacggt aaactgcccc cttggcagta catcaagtgt	840
atcatatgcc aagtccgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt	900
atgcccagta catgacctta cgggactttc ctacttggca gtacatctac gtattagtca	960
tcgctattac catggtgatg cggttttggc agtacaccaa tgggcgtgga tagcggtttg	1020
actcacgggg atttccaagt ctccacccca ttgacgtcaa tgggagtttg ttttggcacc	1080
aaaatcaacg ggactttcca aaatgtcgta ataaccccg cccgttgacg caaatgggcg	1140
gtaggcgtgt acggtgggag gtctatataa gcagagctcg tttagtgaac cgtcagatcg	1200
cctggaacgc catccacgct gttttgacct ccatagaaga cagcgatcg gccaccatgg	1260
tgagcaaggg cgaggagctg ttcaccgggg tggtgcccat cctggtcgag ctggacggcg	1320
acgtaaacgg ccacaagttc agcgtgtccg gcgagggcga gggcgatgcc acctacggca	1380
agctgaccct gaagttcatc tgcaccaccg gcaagctgcc cgtgccctgg cccaccctcg	1440
tgaccaccct gacctacggc gtgcagtgtc tcagccgcta ccccgaccac atgaagcagc	1500
acgacttctt caagtccgcc atgcccgaag gctacgtcca ggagcgcacc atcttcttca	1560
aggacgacgg caactacaag acccgcgccg aggtgaagtt cgagggcgac accctggtga	1620
accgcatcga gctgaagggc atcgacttca aggaggacgg caacatcctg gggcacaagc	1680
tggagtacaa ctacaacagc cacaacgtct atatcatggc cgacaagcag aagaacggca	1740
tcaagtgaa cttcaagatc cgccacaaca tcgaggacgg cagcgtgcag ctcgccgacc	1800
actaccagca gaacaccccc atcggcgacg gccccgtgct gctgcccgc aaccactacc	1860
tgagcaccca gtccgccctg agcaaagacc ccaacgagaa gcgcgatcac atggtcctgc	1920
tggagttcgt gaccgccgcc gggatcactc tcggcatgga cgagctttac aagtagtgag	1980
tttaaactcc catttaaagtg tgagggttaa tgcttcgagc agacatgata agatacattg	2040

atgagtttgg	acaaaccaca	actagaatgc	agtgaaaaaa	atgcttttatt	tgtgaaat	2100
gtgatgctat	tgctttat	gtaaccatta	taagctgcaa	taaacaagtt	aacaacaaca	2160
attgcattca	ttttatgttt	caggttcagg	gggagatgtg	ggaggttttt	taaagcaagt	2220
aaaacctcta	caaagtgtgt	aaaataacta	taacggtcct	aaggtagcga	gtgagtagtg	2280
ttctggggcg	ggggaggacc	tgcatgaggg	ccagaataac	tgaaatctgt	gcttttctgt	2340
gtgttcagc	agcatgagcg	gaagcggctc	ctttgaggga	ggggatttca	gcccttatct	2400
gacggggcgt	ctcccctcct	gggcgggagt	gcgtcagaat	gtgatgggat	ccacggtgga	2460
cggccggccc	gtgcagccc	cgaactcttc	aaccctgacc	tatgcaaccc	tgagctcttc	2520
gtcgttggac	gcagctgcc	ccgcagctgc	tgcatctgcc	gccagcgccg	tgcgcggaat	2580
ggccatgggc	gccggctact	acggcactct	ggtggccaac	tcgagttcca	ccaataatcc	2640
cgccagcctg	aacgaggaga	agctgttgct	gctgatggcc	cagctcgagg	ccttgacca	2700
gcgctgggc	gagctgacc	agcaggtggc	tcagctgcag	gagcagacgc	gggccgcggt	2760
tgccacggtg	aaatccaaat	aaaaaatgaa	tcaataaata	aacggagacg	gttggtgatt	2820
ttaacacaga	gtctgaatct	ttatttgatt	tttcgcgcgc	ggtaggccct	ggaccaccgg	2880
tctcgaatcat	tgagcaccgc	gtggatcttt	tccaggaccc	ggtagagggtg	ggcttggaatg	2940
ttgaggtaca	tgggcatgag	cccgtcccgg	gggtggaggt	agctccattg	cagggcctcg	3000
tgctcggggg	tggtgttgta	aatcaccag	tcatagcagg	ggcgagggc	atggtgttgc	3060
acaatatctt	tgaggaggag	actgatggcc	acgggcagcc	ctttggtgta	ggtgtttaca	3120
aatctgttga	gctgggaggg	atgcatgcgg	ggggagatga	ggtgcatctt	ggcctggatc	3180
ttgagattgg	cgatgttacc	gccagatcc	cgcctggggg	tcatgttggtg	caggaccacc	3240
agcacggtgt	atccggtgca	cttggggaat	ttatcatgca	acttggaagg	gaaggcgtga	3300
aagaatttgg	cgacgccttt	gtgcccgc	aggttttcca	tgcaactcatc	catgatgatg	3360
gcgatgggcc	cgtgggcggc	ggcctgggca	aagacgtttc	gggggtcgga	cacatcatag	3420
ttgtggtcct	gggtgaggtc	atcataggcc	attttaatga	atttggggcg	gaggggtgccg	3480
gactggggga	caaaggtacc	ctcgatccc	ggggcgtagt	tcccctcaca	gatctgcatc	3540
tcccaggctt	tgagctcgga	gggggggatc	atgtccacct	gcggggcgat	aaagaacacg	3600
gtttccgggg	cgggggagat	gagctgggcc	gaaagcaagt	tccggagcag	ctgggacttg	3660

ccgcagccgg tggggccgta gatgaccccg atgaccggct gcaggtggta gttgagggag	3720
agacagctgc cgtcctcccc gagggagggg gccacctcgt tcatcatctc gcgcacgtgc	3780
atgttctcgc gcaccagttc cgccaggagg cgctctcccc ccagggatag gagctcctgg	3840
agcgaggcga agtttttcag cggcttgagt ccgtcggcca tgggcatttt ggagagggtt	3900
tgttgcaaga gttccaggcg gtcccagagc tcggtgatgt gctctacggc atctcgatcc	3960
agcagacctc ctcgtttcgc gggttgggac ggctgcggga gtagggcacc agacgatggg	4020
cgtccagcgc agccagggtc cggctccttc agggctcgag cgtccgcgtc aggggtggctc	4080
ccgtcacggg gaaggggtgc gcgccgggct gggcgcttgc gaggggtgcgc ttcaggctca	4140
tccggctggg cgaaaaccgc tcccgatcgg cgccctgcgc gtcggccagg tagcaattga	4200
ccatgagttc gtagttgagc gcctcggccg cgtggccttt ggcgcgagc ttacctttgg	4260
aagtctgccc gcaggcggga cagaggaggg acttgagggc gtagagcttg ggggcgagga	4320
agacggactc gggggcgtag gcgtccgcgc cgcagtgggc gcagacggtc tcgcactcca	4380
cgagccaggg gaggtcgggc tggtcggggg caaaaaccag tttcccgcg ttccttttga	4440
tgcgtttctt acctttggc tccatgagct cgtgtccccg ctgggtgaca aagaggctgt	4500
ccgtgtcccc gtagaccgac tttatgggcc ggtcctcgag cgggtgtgccg cggtcctcct	4560
cgtagaggaa ccccgccac tccgagacga aagcccgggt ccaggccagc acgaaggagg	4620
ccacgtggga cgggtagcgg tcgttgtcca ccagcgggtc caccttttcc agggtatgca	4680
aacacatgtc cccctcgtcc acatccagga aggtgattgg cttgtaagt taggccacgt	4740
gaccgggggt cccggccggg ggggtataaa aggggtgcggg tccctgctcg tcctcactgt	4800
cttccggatc gctgtccagg agcgccagct gttggggtag gtattccctc tcgaaggcgg	4860
gcatgacctc ggcactcagg ttgtcagttt ctagaaacga ggaggatttg atattgacgg	4920
tgccggcgga gatgcctttc aagagcccct cgtccatctg gtcagaaaag acgatctttt	4980
tgttgctgag cttggtggcg aaggagccgt agagggcgtt ggagaggagc ttggcgatgg	5040
agcgcatggg ctggtttttt tccttgtcgg cgcgtcctt ggcggcgatg ttgagctgca	5100
cgtactcgcg cgccacgcac ttccattcgg ggaagacggg ggtcagctcg tcgggcacga	5160
ttctgacctg ccagccccga ttatgcaggg tgatgagggt cacactgggt gccacctcgc	5220
cgcgcagggg ctcattagtc cagcagaggc gtccgccctt gcgcgagcag aaggggggca	5280

gggggtccag catgacctcg tcgggggggt cggcatcgat ggtgaagatg ccgggcagga	5340
ggtcgggggtc aaagtagctg atggaagtgg ccagatcgtc cagggcagct tgccattcgc	5400
gcacggccag cgcgctctcg tagggactga ggggcgtgcc ccagggcagtg ggatgggtaa	5460
gcgcggaggc gtacatgccg cagatgtcgt agacgtagag gggctcctcg aggatgccga	5520
tgtaggtggg gtagcagcgc cccccgcgga tgctggcgcg cacgtagtca tacagctcgt	5580
gcgagggggc gaggagcccc gggcccaggt tggcgcgact gggcttttcg gcgcggtaga	5640
cgatctggcg gaaaatggca tgcgagttgg aggagatggg gggcctttgg aagatgttga	5700
agtgggcgtg gggcagtcgc accgagtcgc ggatgaagtg ggcgtaggag tcttgacgct	5760
tggcgacgag ctcggcggtg actaggacgt ccagagcgca gtagtcgagg gtctcctgga	5820
tgatgtcata cttgagctgt cccttttggt tccacagctc gcggttgaga aggaactctt	5880
cgcggtcctt ccagtactct tcgaggggga acccgtcctg atctgcacgg taagagccta	5940
gcatgtagaa ctggttgacg gccttgtagg cgcagcagcc cttctccacg gggagggcgt	6000
aggcctgggc ggccttgcg caggaggtgt gcgtgagggc gaaagtgtcc ctgacatga	6060
ccttgaggaa ctggtgcttg aagtcgatat cgtcgcagcc cccctgctcc cagagctgga	6120
agtccgtcg cttctttagt gcgggggttg gcaaagcgaa agtaacatcg ttgaagagga	6180
tcttgcccg gcggggcata aagttgcgag tgatgcggaa aggttggggc acctcgcccc	6240
ggttgttgat gacctgggag gcgagcacga tctcgtcgaa gccgttgatg ttgtggccca	6300
cgatgtagag ttccacgaat cgcggacggc ccttgacgtg gggcagtttc ttgagctcct	6360
cgtaggtgag ctcgtcgggg tcgctgagcc cgtgctgctc gagcgcccag tcggcgagat	6420
gggggttggc gcggaggaag gaagtccaga gatccacggc cagggcgggt tgacagcgt	6480
cccggtactg acggaactgc tgcccagcgg ccattttttc gggggtgacg cagtagaagg	6540
tgcggggggtc cccgtgccag cgatccatt tgagctggag ggcgagatcg agggcgagct	6600
cgacgagccg gtcgtccccg gagagtttca tgaccagcat gaaggggacg agctgcttgc	6660
cgaaggaccc catccaggtg taggtttcca catcgtaggt gaggaagagc ctttcggtgc	6720
gaggatgcga gccgatgggg aagaactgga tctcctgcca ccaattggag gaatggctgt	6780
tgatgtgatg gaagtagaaa tgccgacggc gcgccgaaca ctcgtgcttg tgtttataca	6840
agcggccaca gtgctcgcaa cgctgcacgg gatgcacgtg ctgcacgagc tgtacctgag	6900

ttcctttgac gaggaatttc agtgggaagt ggagtcgtgg cgcctgcatc tcgtgctgta	6960
ctacgtcgtg gtggtcggcc tggccctctt ctgcctcgat ggtggtcgatg ctgacgagcc	7020
cgcgcgggag gcaggtccag acctcggcgc gagcgggtcg gagagcgagg acgagggcgc	7080
gcaggccgga gctgtccagg gtcctgagac gctgcggagt caggtcagtg ggacgcggcg	7140
gcgcgcgggt gacttgcagg agtttttcca gggcgcgcgg gaggtccaga tgggtacttga	7200
tctccaccgc gccattggtg gcgacgtcga tggcttgtag ggtcccgtgc ccctgggggtg	7260
tgaccaccgt cccccgtttc ttcttgggcg gctggggcga cgggggcggg gcctcttcca	7320
tggttagaag cggcggcgag gacgcgcgcc gggcggcagg ggccggctcg ggcccggagg	7380
caggggcggc aggggcacgt cggcgccgcg cgcgggtagg ttctgggtact gcgcccggag	7440
aagactggcg tgagcgacga cgcgacgggt gacgtcctgg atctgacgcc tctgggtgaa	7500
ggccacggga cccgtgagtt tgaacctgaa agagagttcg acagaatcaa tctcggtatc	7560
gttgacggcg gcctgccgca ggatctcttg cacgtcgccc gagttgtcct ggtaggcgat	7620
ctcggtcgat aactgctcga tctcctcttc ttgaaggctt ccgcggccgg cgcgctccac	7680
ggtggccgcg aggtcgttgg agatgcggcc catgagctgc gagaaggcgt tcatgcccgc	7740
ctcgttccag acgcggctgt agaccacgac gccctcggga tcgcgggcgc gcatgaccac	7800
ctgggcgagg ttgagctcca cgtggcgcgt gaagaccgcg tagttgcaga ggcgctggta	7860
gaggtagttg agcgtggtgg cgatgtgctc ggtgacgaag aaatacatga tccagcggcg	7920
gagcggcatc tcgctgacgt cgcgcagcgc ctccaaacgt tccatggcct cgtaaaaagtc	7980
cacggcgaag ttgaaaaact gggagttgcg cgccgagacg gtcaactcct cctccagaag	8040
acggatgagc tcggcgatgg tggcgcgcac ctgcgcctcg aaggcccccg ggagttcctc	8100
cacttcctct tcttcctcct ccactaacat ctcttctact tcctcctcag gcggcagtgg	8160
tggcggggga gggggcctgc gtcgccggcg gcgcacgggc agacggtcga tgaagcgtc	8220
gatggtctcg ccgcgccggc gtcgcatggt ctcggtgacg gcgcgcccgt cctcgcgggg	8280
ccgcagcgtg aagacgccgc cgcgcatctc caggtggccg ggggggtccc cgttgggcag	8340
ggagagggcg ctgacgatgc atcttatcaa ttgccccgta gggactccgc gcaaggacct	8400
gagcgtctcg agatccacgg gatctgaaaa ccgctgaacg aaggcttcga gccagtcgca	8460
gtcgcaaggt aggctgagca cggtttcttc tggcgggtca tgttggttgg gagcggggcg	8520

ggc gatgctg ctggtgatga agttgaaata ggcggttctg agacggcgga tgggtggcgag	8580
gagcaccagg tctttgggcc cggcttgctg gatgcgcaga cggtcggcca tgccccaggc	8640
gtggtcctga cacctggcca ggtccttgta gtagtcctgc atgagccgct ccacgggcac	8700
ctcctcctcg cccgcgcggc cgtgcatgcg cgtgagcccg aagccgcgct ggggctggac	8760
gagcgccagg tcggcgacga cgcgctcggc gaggatggct tgctggatct gggtgagggt	8820
ggtctggaag tcatcaaagt cgacgaagcg gtggtaggct ccggtgttga tgggtgtagga	8880
gcagttggcc atgacggacc agttgacggt ctggtggccc ggacgcacga gctcgtggta	8940
cttgaggcgc gagtaggcgc gcgtgtcgaa gatgtagtcg ttgcagggtgc gcaccaggta	9000
ctggtagccg atgaggaagt gcggcgggcg ctggcggtag agcggccatc gctcggtaggc	9060
ggggcgccg ggcgcgaggc cctcgagcat ggtgcggtgg tagccgtaga tgtacctgga	9120
catccagggtg atgccggcgg cgggtggtgga ggcgcgcggg aactcgcgga cgcggttcca	9180
gatgttgcg agcggcagga agtagttcat ggtgggcacg gtctggcccg tgaggcgcgc	9240
gcagtcgtgg atgctctata cgggcaaaaa cgaaagcggc cagcggctcg actccgtggc	9300
ctggaggcta agcgaacggg ttgggctgcg cgtgtacccc ggttcgaatc tcgaatcagg	9360
ctggagccgc agctaactg gtattggcac tcccgtctcg acccaagcct gcaccaacct	9420
tccaggatac ggaggcgggt cgttttgcaa cttttttttg gaggccggat gagactagta	9480
agcgcggaaa gcggccgacc gcgatggctc gctgccgtag tctggagaag aatcgccagg	9540
gttgcgttgc ggtgtgcccc ggttcgaggc cggccggatt ccgcggctaa cgagggcgtg	9600
gctgccccgt cgtttccaag accccatagc cagccgactt ctccagttac ggagcgagcc	9660
cctcttttgt tttgtttgtt ttgtccagat gcatcccgtg ctgcggcaga tgcgccccca	9720
ccaccctcca ccgcaacaac agccccctcc acagccggcg cttctgcccc cgccccagca	9780
gcaacttcca gccacgaccg ccgcggccgc cgtgagcggg gctggacaga gttatgatca	9840
ccagctggcc ttggaagagg gcgaggggct ggcgcgccctg ggggcgtcgt cgccggagcg	9900
gcacccgcgc gtgcagatga aaagggacgc tcgcgaggcc tacgtgcca agcagaacct	9960
gttcagagac aggagcggcg aggagcccga ggagatgcgc gcggcccggg tccacgcggg	10020
gcgggagctg cggcgcggcc tggaccgaaa gagggtgctg agggacgagg atttcgaggc	10080
ggacgagctg acggggatca gccccgcgcg cgcgcacgtg gccgcggcca acctggtcac	10140

ggcgtacgag	cagaccgtga	aggaggagag	caacttccaa	aaatccttca	acaaccacgt	10200
gcgcaccttg	atcgcgcgcg	aggaggtgac	cctgggcctg	atgcacctgt	gggacctgct	10260
ggaggccatc	gtgcagaacc	ccaccagcaa	gccgctgacg	gcgcagctgt	tcctggtggt	10320
gcagcatagt	cgggacaacg	aagcgttcag	ggaggcgctg	ctgaatatca	ccgagcccga	10380
gggccgctgg	ctcctggacc	tgggtgaacat	tctgcagagc	atcgtggtgc	aggagcgcg	10440
gctgccgctg	tccgagaagc	tggcggccat	caacttctcg	gtgctgagtt	tgggcaagta	10500
ctacgctagg	aagatctaca	agaccccgta	cgtgcccata	gacaaggagg	tgaagatcga	10560
cgggttttac	atcgcatga	ccctgaaagt	gctgaccctg	agcgacgatc	tgggggtgta	10620
ccgcaacgac	aggatgcacc	gtgcggtgag	cgccagcagg	cggcgcgagc	tgagcgacca	10680
ggagctgatg	catagtctgc	agcgggccct	gaccggggcc	gggaccgagg	gggagagcta	10740
ctttgacatg	ggcgcgacc	tgcactggca	gcccagccgc	cgggccttgg	aggcggcggc	10800
aggaccctac	gtagaagagg	tggacgatga	ggtggacgag	gagggcgagt	acctggaaga	10860
ctgatggcgc	gaccgtat	ttgctagatg	caacaacaac	agccacctcc	tgatcccgcg	10920
atcgggcg	cgctgcagag	ccagccgtcc	ggcattaact	cctcggacga	ttggaccag	10980
gccatgcaac	gcatcatggc	gctgacgacc	cgcaaccccg	aagcctttag	acagcagccc	11040
caggccaacc	ggctctcg	catcctggag	gccgtggtgc	cctcgcgctc	caacccacg	11100
cacgagaagg	tcctggccat	cgtgaacg	ctggtggaga	acaaggccat	ccgcggcgac	11160
gaggccggcc	tgggtgtaaa	cgcgctgctg	gagcgcggtg	cccgctacaa	cagcaccaac	11220
gtgcagacca	acctggaccg	catggtgacc	gacgtgcgcg	aggccgtggc	ccagcgcgag	11280
cggttccacc	gcgagtccaa	cctgggatcc	atggtggcgc	tgaacgcctt	cctcagcacc	11340
cagcccgcca	acgtgccccg	gggccaggag	gactacacca	acttcatcag	cgccctgcgc	11400
ctgatggtga	ccgaggtgcc	ccagagcgag	gtgtaccagt	ccgggccgga	ctacttcttc	11460
cagaccagtc	gccagggctt	gcagaccgtg	aacctgagcc	aggctttcaa	gaacttgag	11520
ggcctgtggg	gcgtgcaggc	cccggtcggg	gaccgcgcga	cgggtgtcag	cctgctgacg	11580
ccgaactcgc	gcctgctgct	gctgctggtg	gcccccttca	cggacagcgg	cagcatcaac	11640
cgcaactcgt	acctgggcta	cctgattaac	ctgtaccgcg	aggccatcgg	ccaggcgcac	11700
gtggacgagc	agacctacca	ggagatcacc	cacgtgagcc	gcgccctggg	ccaggacgac	11760

ccgggcaacc	tggaagccac	cctgaacttt	ttgctgacca	accggtcgca	gaagatcccc	11820
ccccagtagc	cgctcagcac	cgaggaggag	cgcatcctgc	gttacgtgca	gcagagcgtg	11880
ggcctgttcc	tgatgcagga	ggggggccacc	cccagcgccg	cgctcgacat	gaccgcgcgc	11940
aacatggagc	ccagcatgta	cgccagcaac	cgcccgttca	tcaataaact	gatggactac	12000
ttgcatcggg	cggccgcat	gaactctgac	tatttcacca	acgccatcct	gaatccccac	12060
tggctccgc	cgccgggggtt	ctacacgggc	gagtacgaca	tgcccgaccc	caatgacggg	12120
ttcctgtggg	acgatgtgga	cagcagcgtg	ttctcccccc	gaccgggtgc	taacgagcgc	12180
cccttgtgga	agaaggaagg	cagcgaccga	cgcccgctct	cggcgctgtc	cggccgcgag	12240
ggtgctgccg	cggcgggtgcc	cgaggccgcc	agtcctttcc	cgagcttgcc	cttctcgctg	12300
aacagtatcc	gcagcagcga	gctgggcagg	atcacgcgcc	cgcgcttgct	gggcgaagag	12360
gagtacttga	atgactcgct	gttgagaccc	gagcgggaga	agaacttccc	caataacggg	12420
atagaaagcc	tggtggacaa	gatgagccgc	tggaagacgt	atgcgcagga	gcacagggac	12480
gatccccggg	cgtcgcaggg	ggccacgagc	cggggcagcg	ccgcccgtaa	acgccggtgg	12540
cacgacaggc	agcgggggaca	gatgtgggac	gatgaggact	ccgccgacga	cagcagcgtg	12600
ttggacttgg	gtgggagtgg	taacccgttc	gctcacctgc	gccccgtat	cgggcgcatg	12660
atgtaagaga	aaccgaaaat	aaatgatact	caccaaggcc	atggcgacca	gcgtgcgttc	12720
gtttcttctc	tgttgttggt	gtatctagta	tgatgaggcg	tgcgtaaccg	gagggtcctc	12780
ctccctcgta	cgagagcgtg	atgcagcagg	cgatggcggc	ggcggcgatg	cagccccgc	12840
tggaggctcc	ttacgtgccc	ccgcggtacc	tggcgcctac	ggagggggcgg	aacagcattc	12900
gttactcgga	gctggcacc	ttgtacgata	ccaccgggtt	gtacctgggtg	gacaacaagt	12960
cggcggacat	cgctcgctg	aactaccaga	acgaccacag	caacttcctg	accaccgtgg	13020
tgcagaacaa	tgacttcacc	cccacggagg	ccagcaccca	gaccatcaac	tttgacgagc	13080
gctcgcggtg	gggcggccag	ctgaaaacca	tcatgcacac	caacatgccc	aacgtgaacg	13140
agttcatgta	cagcaacaag	ttcaaggcgc	gggtgatggg	ctcccgcaag	acccccaatg	13200
gggtgacagt	gacagaggat	tatgatggta	gtcaggatga	gctgaagtat	gaatgggtgg	13260
aatttgagct	gcccgaaggc	aacttctcgg	tgaccatgac	catcgacctg	atgaacaacg	13320
ccatcatcga	caattacttg	gcggtggggc	ggcagaacgg	ggtgctggag	agcgacatcg	13380

gcgtgaagtt	cgacactagg	aacttcaggc	tgggctggga	ccccgtgacc	gagctgggtca	13440
tgtccgggggt	gtacaccaac	gaggctttcc	atccccgatat	tgtcttgctg	cccgggtgcg	13500
gggtggactt	caccgagagc	cgcctcagca	acctgctggg	cattcgcaag	aggcagccct	13560
tccaggaagg	cttccagatc	atgtacgagg	atctggaggg	gggcaacatc	cccgcgctcc	13620
tggatgtcga	cgcctatgag	aaaagcaagg	aggatgcagc	agctgaagca	actgcagccg	13680
tagctaccgc	ctctaccgag	gtcaggggcg	ataattttgc	aagcgccgca	gcagtggcag	13740
cggccgaggc	ggctgaaacc	gaaagtaaga	tagtcattca	gccgggtggag	aaggatagca	13800
agaacaggag	ctacaacgta	ctaccggaca	agataaacac	cgcctaccgc	agctgggtacc	13860
tagcctacaa	ctatggcgac	cccgagaagg	gcgtgcgctc	ctggacgctg	ctcaccacct	13920
cggacgtcac	ctgcggcggtg	gagcaagtct	actggtcgct	gcccgcacatg	atgcaagacc	13980
cggtcacctt	ccgctccacg	cgtcaagtta	gcaactaccc	ggtgggtgggc	gccgagctcc	14040
tgtccgtcta	ctccaagagc	ttcttcaacg	agcaggccgt	ctactcgag	cagctgcgcg	14100
ccttcacctc	gcttacgcac	gtcttcaacc	gcttccccga	gaaccagatc	ctcgtccgcc	14160
cgtccgcgcc	caccattacc	accgtcagtg	aaaacgttcc	tgctctcaca	gatcacggga	14220
ccctgccgct	gcgcagcagt	atccggggag	tccagcgcg	gaccgttact	gacgccagac	14280
gccgcacctg	cccctacgtc	tacaaggccc	tgggcatagt	cgcgccgcgc	gtcctctcga	14340
gccgcacctt	ctaaatgtcc	atttctatct	cgcccagtaa	taacaccggt	tggggcctgc	14400
gcgcgcccag	caagatgtac	ggaggcgctc	gccaacgctc	cacgcaacac	cccgtgcgcg	14460
tgcgcgggca	cttccgcgct	ccctggggcg	ccctcaaggg	ccgcgtgcgg	tcgcgcacca	14520
ccgtcgacga	cgtgatcgac	caggtgggtg	ccgacgcgcg	caactacacc	cccgccgccg	14580
cgtccgtctc	caccgtggac	gccgtcatcg	acagcgtggt	ggccgacgcg	cgccggtacg	14640
cccgcgccaa	gagccggcgg	cggcgcatcg	cccggcgggca	ccggagcacc	cccgccatgc	14700
gcgcggcgcg	agccttgctg	cgcagggcca	ggcgcacggg	acgcagggcc	atgctcaggg	14760
cggccagacg	cgcggcttca	ggcgccagcg	ccggcaggac	ccggagacgc	gcggccacgg	14820
cggcgggcagc	ggccatcgcc	agcatgtccc	gcccgcggcg	agggaacgtg	tactgggtgc	14880
gcgacgccgc	caccggtgtg	cgcgtgcccg	tgcgcacccg	ccccctcgc	acttgaagat	14940
gttcacttcg	cgatgttgat	gtgtcccagc	ggcgaggagg	atgtccaagc	gcaaattcaa	15000

ggaagagatg	ctccaggtca	tcgcgctga	gatctacggc	cctgcggtgg	tgaaggagga	15060
aagaaagccc	cgcaaaatca	agcgggtcaa	aaaggacaaa	aaggaagaag	aaagtgatgt	15120
ggacggattg	gtggagtttg	tgcgcgagtt	cgcccccg	cggcgcgtgc	agtggcgcgg	15180
gcggaaggtg	caaccggtgc	tgagaccg	caccaccgtg	gtcttcacgc	ccggcgagcg	15240
ctccggcacc	gcttccaagc	gctcctacga	cgagggtgtac	ggggatgatg	atattctgga	15300
gcaggcggcc	gagcgcctgg	gcgagtttgc	ttacggcaag	cgagccggtt	ccgcaccgaa	15360
ggaagaggcg	gtgtccatcc	cgctggacca	cggcaacccc	acgccgagcc	tcaagcccgt	15420
gaccttgag	caggtgctgc	cgaccgcggc	gccgcgccgg	gggttcaagc	gcgagggcga	15480
ggatctgtac	cccaccatgc	agctgatggt	gcccgaagcgc	cagaagctgg	aagacgtgct	15540
ggagaccatg	aaggtggacc	cggacgtgca	gcccgaagtc	aaggtgcggc	ccatcaagca	15600
ggtggccccg	ggcctggg	tgagaccgt	ggacatcaag	attcccacgg	agcccatgga	15660
aacgcagacc	gagcccatga	tcaagcccag	caccagcacc	atggaggtgc	agacggatcc	15720
ctggatgcca	tcggctccta	gtcgaagacc	ccggcgcaag	tacggcgcg	ccagcctgct	15780
gatgccaac	tacgcgctgc	atccttccat	catccccacg	ccgggctacc	gcggcacgcg	15840
cttctaccgc	ggtcatacca	gcagccgccg	ccgcaagacc	accactcgcc	gccgccgtcg	15900
ccgcaccgcc	gctgcaacca	cccctgccgc	cctggtgcgg	agagtgtacc	gccgcggccg	15960
cgcacctctg	accctgccgc	gcgcgcgcta	ccaccgagc	atcgccat	aaactttcgc	16020
ctgctttgca	gatcaatggc	cctcacatgc	cgcttcgcg	ttcccattac	gggctaccga	16080
ggaagaaaac	cgcgccgtag	aaggctggcg	gggaacggga	tgcgtcgcca	ccaccaccgg	16140
cggcggcgcg	ccatcagcaa	gcggttgggg	ggaggcttcc	tgcccgcgct	gatccccatc	16200
atcgccgcgg	cgatcggggc	gatccccggc	attgcttccg	tggcgggtgca	ggcctctcag	16260
cgccactgag	acacacttgg	aaacatcttg	taataaacca	atggactctg	acgctcctgg	16320
tcctgtgatg	tgttttcgta	gacagatgga	agacatcaat	ttttcgcccc	tggctccgcg	16380
acacggcacg	cggccgttca	tgggcacctg	gagcgacatc	ggcaccagcc	aactgaacgg	16440
gggcgccttc	aattggagca	gtctctggag	cgggcttaag	aatttcgggt	ccacgcttaa	16500
aacctatggc	agcaaggcgt	ggaacagcac	cacagggcag	gcgctgaggg	ataagctgaa	16560
agagcagaac	ttccagcaga	aggtggtcga	tgggctcgcc	tcgggcatca	acggggtgggt	16620

ggacctggcc	aaccaggccg	tgcagcggca	gatcaacagc	cgcctggacc	cggtgccgcc	16680
cgccggctcc	gtggagatgc	cgcaggtgga	ggaggagctg	cctcccctgg	acaagcgggg	16740
cgagaagcga	ccccgccccg	atgcggagga	gacgctgctg	acgcacacgg	acgagccgcc	16800
cccgtacgag	gaggcggtga	aactgggtct	gcccaccacg	cggcccatcg	cgcccctggc	16860
caccggggtg	ctgaaacccg	aaaagcccgc	gaccctggac	ttgcctcctc	cccagccttc	16920
ccgcccctct	acagtggcta	agcccctgcc	gccggtggcc	gtggcccgcg	cgcgaccggg	16980
gggcaccgcc	cgccctcatg	cgaactggca	gagcactctg	aacagcatcg	tgggtctggg	17040
agtgcagagt	gtgaagcgcc	gccgctgcta	ttaaacctac	cgtagcgctt	aacttgcttg	17100
tctgtgtgtg	tatgtattat	gtcgccgccg	ccgctgtcca	ccagaaggag	gagtgaagag	17160
gcgcgtcgcc	gagttgcaag	atggccaccc	catcgatgct	gccccagtgg	gcgtacatgc	17220
acatcgccgg	acaggacgct	tcggagtacc	tgagtccggg	tctggtgcag	tttgcccgcg	17280
ccacagacac	ctacttcagt	ctggggaaca	agtttaggaa	ccccacgggtg	gcgcccacgc	17340
acgatgtgac	caccgaccgc	agccagcggc	tgacgctgcg	cttcgtgccc	gtggaccgcg	17400
aggacaacac	ctactcgtac	aaagtgcgct	acacgctggc	cgtgggcgac	aaccgcgtgc	17460
tggacatggc	cagcacctac	tttgacatcc	gcggcgtgct	ggatcggggc	cctagcttca	17520
aaccctactc	cggcaccgcc	tacaacagtc	tggcccccaa	gggagcaccc	aacacttgtc	17580
agtggacata	taaagccgat	ggtgaaactg	ccacagaaaa	aacctataca	tatggaaatg	17640
cacccgtgca	gggcattaac	atcacaaaag	atgggtattca	acttggaact	gacaccgatg	17700
atcagccaat	ctacgcagat	aaaacctatc	agcctgaacc	tcaagtgggt	gatgctgaat	17760
ggcatgacat	cactggtact	gatgaaaagt	atggaggcag	agctcttaag	cctgatacca	17820
aaatgaagcc	ttgttatggt	tcttttgcca	agcctactaa	taaagaagga	ggtcaggcaa	17880
atgtgaaaac	aggaacaggc	actactaaag	aatatgacat	agacatggct	ttctttgaca	17940
acagaagtgc	ggctgctgct	ggcctagctc	cagaaattgt	tttgataact	gaaaatgtgg	18000
atttggaaac	tccagatacc	catattgtat	acaaagcagg	cacagatgac	agcagctctt	18060
ctattaattt	gggtcagcaa	gccatgcccc	acagacctaa	ctacattggt	ttcagagaca	18120
actttatcgg	gctcatgtac	tacaacagca	ctggcaatat	gggggtgctg	gccggtcagg	18180
cttctcagct	gaatgctgtg	gttgacttgc	aagacagaaa	caccgagctg	tcctaccagc	18240

tcttgcttga	ctctctgggt	gacagaaccc	ggtatttcag	tatgtggaat	caggcgggtg	18300
acagctatga	tcctgatgtg	cgcattattg	aaaatcatgg	tgtggaggat	gaacttccca	18360
actattgttt	ccctctggat	gctgttggca	gaacagatac	ttatcaggga	attaaggcta	18420
atggaactga	tcaaaccaca	tggaccaaag	atgacagtgt	caatgatgct	aatgagatag	18480
gcaagggtaa	tccattcgcc	atggaaatca	acatccaagc	caacctgtgg	aggaacttcc	18540
tctacgcaa	cgtggccctg	tacctgccc	actcttaca	gtacacgccg	gccaatgtta	18600
ccctgcccac	caacaccaac	acctacgatt	acatgaacgg	ccgggtgggtg	gcgccctcgc	18660
tggtggactc	ctacatcaac	atcggggcgc	gctggtcgct	ggatcccatg	gacaacgtga	18720
accccttcaa	ccaccaccgc	aatgcggggc	tgcgctaccg	ctccatgctc	ctgggcaacg	18780
ggcgctacgt	gcccttcac	atccaggtgc	cccagaaatt	tttcgccatc	aagagcctcc	18840
tgctcctgcc	cgggtcctac	acctacgagt	ggaacttccg	caaggacgtc	aacatgatcc	18900
tgagagctc	cctcggcaac	gacctgcgca	cggacggggc	ctccatctcc	ttcaccagca	18960
tcaacctcta	cgccaccttc	ttccccatgg	cgcacaacac	ggcctccacg	ctcgaggcca	19020
tgctgcgcaa	cgacaccaac	gaccagtcct	tcaacgacta	cctctcggcg	gccaacatgc	19080
tctaccccat	cccggccaac	gccaccaacg	tgcccatctc	catcccctcg	cgcaactggg	19140
ccgccttccg	cggctgggtc	ttcacgcgtc	tcaagaccaa	ggagacgccc	tcgctgggct	19200
ccgggttcga	cccctacttc	gtctactcgg	gtcccatccc	ctacctcgac	ggcaccttct	19260
acctcaacca	caccttcaag	aaggtctcca	tcaccttcga	ctcctccgtc	agctggccccg	19320
gcaacgaccg	gctcctgacg	cccaacgagt	tcgaaatcaa	gcgcaccgtc	gacggcgagg	19380
gctacaacgt	ggcccagtg	aacatgacca	aggactgggt	cctgggtccag	atgctggccc	19440
actacaacat	cggctaccag	ggcttctacg	tgcccagggg	ctacaaggac	cgcatgtact	19500
ccttcttccg	caacttccag	cccatgagcc	gccaggtgggt	ggacgagggtc	aactacaagg	19560
actaccaggc	cgtcaccctg	gcctaccagc	acaacaactc	gggcttcgtc	ggctacctcg	19620
cgcccaccat	gcgccagggc	cagccctacc	ccgccaacta	cccctacccg	ctcatcggca	19680
agagcgccgt	caccagcgtc	accagaaaa	agttcctctg	cgacagggtc	atgtggcgca	19740
tccccttctc	cagcaacttc	atgtccatgg	gcgcgctcac	cgacctcggc	cagaacatgc	19800
tctatgcaa	ctccgcccac	gcgctagaca	tgaatttcga	agtcgacccc	atggatgagt	19860

ccacccttct	ctatgttgtc	ttcgaagtct	tcgacgtcgt	ccgagtgcac	cagccccacc	19920
gcggcgatcat	cgaggccgtc	tacctgcgca	cccccttctc	ggccggtaac	gccaccacct	19980
aagctcttgc	ttcttgcaag	ccatggccgc	gggctccggc	gagcaggagc	tcagggccat	20040
catccgcgac	ctgggctgcg	ggccctactt	cctgggcacc	ttcgataagc	gcttcccggg	20100
attcatggcc	ccgcacaagc	tggcctgcgc	catcgtcaac	acggccggcc	gcgagaccgg	20160
gggcgagcac	tggctggcct	tcgcctggaa	cccgcgctcg	aacacctgct	acctcttcga	20220
ccccctcggg	ttctcggacg	agcgcctcaa	gcagatctac	cagttcgagt	acgagggcct	20280
gctgcgccgc	agcgccctgg	ccaccgagga	ccgctgcgtc	accctggaaa	agtccacca	20340
gaccgtgcag	ggtccgcgct	cggccgcctg	cgggctcttc	tgctgcatgt	tcctgcacgc	20400
cttcgtgcac	tggcccgacc	gccccatgga	caagaacccc	accatgaact	tgctgacggg	20460
ggtgccaac	ggcatgctcc	agtcgcccc	ggtggaaccc	accctgcgcc	gcaaccagga	20520
ggcgctctac	cgcttcctca	actcccactc	cgcctacttt	cgctcccacc	gcgcgcgcat	20580
cgagaaggcc	accgccttcg	accgcatgaa	tcaagacatg	taaaccgtgt	gtgtatgtta	20640
aatgtcttta	ataaacagca	ctttcatgtt	acacatgcat	ctgagatgat	ttatttagaa	20700
atcgaaaggg	ttctgccggg	tctcggcatg	gcccgcgggc	agggacacgt	tgcggaactg	20760
gtacttggcc	agccacttga	actcggggat	cagcagtttg	ggcagcgggg	tgctggggaa	20820
ggagtccgtc	cacagcttcc	gcgtcagttg	cagggcgccc	agcaggtcgg	gcgcggagat	20880
cttgaaatcg	cagttgggac	ccgcgttctg	cgcgcgggag	ttgcggtaca	cggggttgca	20940
gcactggaac	accatcaggg	ccgggtgctt	cacgctcgcc	agcaccgtcg	cgctcggatg	21000
gctctccacg	tcgaggtcct	cggcgttggc	catcccgaag	ggggtcacat	tgacaggtctg	21060
ccttcccatg	gtgggcacgc	acccgggctt	gtggttgcaa	tcgcagtgca	gggggatcag	21120
catcatctgg	gcctgggtcg	cgttcatccc	cgggtacatg	gccttcatga	aagcctccaa	21180
ttgcctgaac	gcctgctggg	ccttggctcc	ctcgggtgaag	aagacccgcg	aggacttgct	21240
agagaactgg	ttgggtggcg	acccggcgct	gtgcacgcag	cagcgcgcg	cgttggtggc	21300
cagctgcacc	acgctgcgcc	cccagcggtt	ctgggtgatc	ttggcccgg	cggggttctc	21360
cttcagcgcg	cgctgcccgt	tctcgctcgc	cacatccatc	tcgatcatgt	gctccttctg	21420
gatcatggtg	gtcccgtgca	ggcaccgcag	cttgccctcg	gcctcgggtg	acccgtgcag	21480

ccacagcgcg caccggtgc actcccagtt cttgtgggcg atctgggaat gcgcgtgcac	21540
gaagccctgc aggaagcggc ccatcatggt ggtcagggtc ttgttgctag tgaaggctcag	21600
cggaatgccg cggtgctcct cgttgatgta cagggtggcag atgcggcggc acacctcgcc	21660
ctgctcgggc atcagctgga agttggcttt caggctcggc tccacgcggc agcgggtccat	21720
cagcatagtc atgattttcca taccctttctc ccaggccgag acgatgggca ggctcatagg	21780
gttcttcacc atcatcttag cgctagcagc cgcggccagg gggtcgctct cgtccagggt	21840
ctcaaagctc cgcttgccgt ctttctcggt gatccgcacc ggggggtagc tgaagcccac	21900
ggccgccagc tcctcctcgg cctgtctttc gtcctcgctg tcctggctga cgtcctgcag	21960
gaccacatgc ttggtcttgc ggggttttctt cttgggcggc agcggcggcg gagatgttgg	22020
agatggcgag ggggagcgcg agttctcgct caccactact atctcttcct cttcttggtc	22080
cgaggccacg cggcggtagg tatgtctctt cgggggcaga ggcgaggcg acgggctctc	22140
gccgccgca cttggcggat ggctggcaga gccccttcg cgttcggggg tgcgctcccg	22200
gcggcgctct gactgacttc ctccgcggcc ggccattgtg ttctcctagg gaggaacaac	22260
aagcatggag actcagccat cgccaacctc gccatctgcc cccaccgccg acgagaagca	22320
gcagcagcag aatgaaagct taaccgcccc gccgcccagc cccgccacct ccgacgcggc	22380
cgtcccagac atgcaagaga tggaggaatc catcgagatt gacctgggct atgtgacgcc	22440
cgcgggagcac gaggaggagc tggcagtgcg cttttcaca gaagagatac accaagaaca	22500
gccagagcag gaagcagaga atgagcagag tcaggctggg ctcgagcatg acggcgacta	22560
cctccacctg agcggggggg aggacgcgct catcaagcat ctggcccggc aggccaccat	22620
cgtcaaggat gcgctgctcg accgcaccga ggtgcccctc agcgtggagg agctcagccg	22680
cgcctacgag ttgaacctct tctcgccgcg cgtgcccccc aagcgccagc ccaatggcac	22740
ctgcgagccc aaccgcgcc tcaacttcta cccggtcttc gcggtgcccg aggccctggc	22800
cacctaccac atctttttca agaaccaaaa gatccccgtc tcctgccgcg ccaaccgcac	22860
ccgcgccgac gcccttttca acctgggtcc cggcgccgc ctacctgata tcgcctcctt	22920
ggaagagggt cccaagatct tcgagggtct gggcagcgac gagactcggg ccgcgaacgc	22980
tctgcaagga gaaggaggag agcatgagca ccacagcgcc ctggtcgagt tggaaggcga	23040
caacgcgcgg ctggcgggtgc taaaacgcac ggtcgagctg acccatttcg cctaccggc	23100

tctgaacctg	ccccccaaag	tcatgagcgc	ggatcatggac	caggtgctca	tcaagcgcgc	23160
gtcgcccatc	tccgaggacg	agggcatgca	agactccgag	gagggaagc	ccgtgggtcag	23220
cgacgagcag	ctggcccggg	ggctgggtcc	taatgctagt	ccccagagtt	tggaagagcg	23280
gcgcaaactc	atgatggccg	tggtcctggg	gaccgtggag	ctggagtgcc	tgcgccgctt	23340
cttcgccgac	gcggagaccc	tgcgcaagg	cgaggagaac	ctgcactacc	tcttcaggca	23400
cgggttcgtg	cgccaggcct	gcaagatctc	caacgtggag	ctgaccaacc	tggtctccta	23460
catgggcatc	ttgcacgaga	accgcctggg	gcagaacgtg	ctgcacacca	ccctgcgcgg	23520
ggaggcccgg	cgcgactaca	tccgcgactg	cgtctacctc	tacctctgcc	acacctggca	23580
gacgggcatg	ggcgtgtggc	agcagtgtct	ggaggagcag	aacctgaaag	agctctgcaa	23640
gctcctgcag	aagaacctca	agggctctgtg	gaccgggttc	gacgagcgca	ccaccgcctc	23700
ggacctggcc	gacctcattt	tccccgagcg	cctcaggctg	acgctgcgca	acggcctgcc	23760
cgactttatg	agccaaagca	tgttgcaaaa	ctttcgctct	ttcatcctcg	aacgctccgg	23820
aatcctgccc	gccacctgct	ccgcgctgcc	ctcggacttc	gtgccgctga	ccttccgcga	23880
gtgccccccg	ccgctgtgga	gccactgcta	cctgctgcgc	ctggccaact	acctggccta	23940
ccactcggac	gtgatcgagg	acgtcagcgg	cgagggcctg	ctcgagtgcc	actgccgctg	24000
caacctctgc	acgccgcacc	gtccctggc	ctgcaacccc	cagctgctga	gcgagacca	24060
gatcatcggc	accttcgagt	tgcaagggcc	cagcgaaggc	gagggttcag	ccgccaaggg	24120
gggtctgaaa	ctcaccgccg	ggctgtggac	ctcggcctac	ttgcgcaagt	tcgtgcccga	24180
ggactaccat	cccttcgaga	tcaggttcta	cgaggaccaa	tcccatccgc	ccaaggccga	24240
gctgtcggcc	tgcgtcatca	cccagggggc	gatcctggcc	caattgcaag	ccatccagaa	24300
atcccgccaa	gaattcttgc	tgaaaaaggg	ccgcgggggtc	tacctcgacc	cccagaccgg	24360
tgaggagctc	aaccccggt	tccccagga	tgccccgagg	aaacaagaag	ctgaaagtgg	24420
agctgccgcc	cgtggaggat	ttggaggaag	actgggagaa	cagcagtcag	gcagaggagg	24480
aggagatgga	ggaagactgg	gacagcactc	aggcagagga	ggacagcctg	caagacagtc	24540
tggaggaaga	cgaggaggag	gcagaggagg	aggtggaaga	agcagccgcc	gccagaccgt	24600
cgtcctcggc	gggggagaaa	gcaagcagca	cggataccat	ctccgctccg	ggtcgggggtc	24660
ccgctcgacc	acacagtaga	tgggacgaga	ccggacgatt	cccgaacccc	accacccaga	24720

ccggttaagaa	ggagcggcag	ggatacaagt	cctggcgggg	gcacaaaaac	gccatcgtct	24780
cctgcttgca	ggcctgcggg	ggcaacatct	ccttcacccg	gcgtacctg	ctcttccacc	24840
gcgggggtgaa	ctttccccgc	aacatcttgc	attactaccg	tcacctccac	agcccctact	24900
acttccaaga	agaggcagca	gcagcagaaa	aagaccagca	gaaaaccagc	agctagaaaa	24960
tccacagcgg	cggcagcagg	tggactgagg	atcgcgggcga	acgagccggc	gcaaaccg	25020
gagctgagga	accggatctt	ttccaccctc	tatgccatct	tccagcagag	tcgggggcag	25080
gagcaggaac	tgaaagtcaa	gaaccgttct	ctgcgctcgc	tcacccgcag	ttgtctgtat	25140
cacaagagcg	aagaccaact	tcagcgcact	ctcgaggacg	ccgaggctct	cttcaacaag	25200
tactgcgcgc	tcactcttaa	agagtagccc	gcgcccggcc	agtcgcagaa	aaaggcggga	25260
attacgtcac	ctgtgccctt	cgccctagcc	gcctccaccc	atcatcatga	gcaaagagat	25320
ttccacgcct	tacatgtgga	gctaccagcc	ccagatgggc	ctggccgccg	gtgccgcca	25380
ggactactcc	acccgcatga	attggctcag	cgccggggcc	gcgatgatct	cacgggtgaa	25440
tgacatccgc	gcccaccgaa	accagatact	cctagaacag	tcagcgctca	ccgccacgcc	25500
ccgcaatcac	ctcaatccgc	gtaattggcc	cgccgccctg	gtgtaccagg	aaattcccca	25560
gcccacgacc	gtactacttc	cgcgagacgc	ccaggccgaa	gtccagctga	ctaactcagg	25620
tgtccagctg	gcgggcggcg	ccaccctgtg	tcgtcaccgc	cccgctcagg	gtataaagcg	25680
gctggtgatc	cggggcagag	gcacacagct	caacgacgag	gtggtgagct	cttcgctggg	25740
tctgcgacct	gacggagtct	tccaactcgc	cggatcgggg	agatcttcct	tcacgcctcg	25800
tcaggccgtc	ctgactttgg	agagttcgtc	ctcgcagccc	cgctcgggtg	gcatcggcac	25860
tctccagttc	gtggaggagt	tcactccctc	ggctctacttc	aacccttct	ccggctcccc	25920
cggccactac	ccggacgagt	tcatcccgaa	cttcgacgcc	atcagcgagt	cgggtggacgg	25980
ctacgattga	atgtcccatg	gtggcgcagc	tgacctagct	cggcttcgac	acctggacca	26040
ctgccgccgc	ttccgctgct	tcgctcggga	tctcgccgag	tttgcttact	ttgagctgcc	26100
cgaggagcac	cctcaggggc	cggcccacgg	agtgcggatc	gtcgtcgaag	ggggcctcga	26160
ctccacactg	cttcggatct	tcagccagcg	tccgatcctg	gtcgagcgcg	agcaaggaca	26220
gacccttctg	actctgtact	gcatctgcaa	ccacccggc	ctgcatgaaa	gtctttgttg	26280
tctgctgtgt	actgagtata	ataaaagctg	agatcagcga	ctactccgga	cttccgtgtg	26340

ttcctgaatc catcaaccag tctttgttct tcaccgggaa cgagaccgag ctccagctcc	26400
agtgtagcc ccacaagaag tacctcacct ggctgttcca gggctccccg atcgccgttg	26460
tcaaccactg cgacaacgac ggagtcctgc tgagcggccc tgccaacctt actttttcca	26520
cccgagaag caagctccag ctcttccaac ccttcctccc cgggacctat cagtgcgtct	26580
cgggaccctg ccatcacacc ttccacctga tcccgaatac cacagcgtcg ctccccgcta	26640
ctaacaacca aactaacctc caccaacgcc accgtcgcga cggccacaat acatgccccat	26700
attagactat gaggccgagc cacagcgacc catgctcccc gctattagtt acttcaatct	26760
aaccggcgga gatgactgac ccactggcca acaacaacgt caacgacctt ctcctggaca	26820
tggacggccg cgcctcggag cagcgactcg cccaacttcg cattcgccag cagcaggaga	26880
gagccgtcaa ggagctgcag gatgcggtgg ccatccacca gtgcaagaga ggcatcttct	26940
gcctggtgaa acaggccaag atctcctacg aggtcactcc aaacgaccat cgcctctcct	27000
acgagctcct gcagcagcgc cagaagttca cctgcctggc cggagtcaac cccatcgtca	27060
tcaccagca gtctggcgat accaaggggt gcatccactg ctcctgcgac tccccgact	27120
gcgtccacac tctgatcaag accctctgcg gcctccgcga cctcctcccc atgaactaat	27180
cacccccta tccagtgaat taaagatcat attgatgatg attttacaga aataaaaaat	27240
aatcatttga tttgaaataa agatacaatc atattgatga tttgagttta acaaaaaaat	27300
aaagaatcac ttacttgaaa tctgatacca ggtctctgtc catgttttct gccaacacca	27360
cttcactccc ctcttcccag ctctggtact gcaggccccg gcgggctgca aacttcctcc	27420
acacgctgaa ggggatgtca aattcctcct gtccctcaat cttcatttta tcttctatca	27480
gatgtccaaa aagcgcgtcc ggggtggatga tgacttcgac cccgtctacc cctacgatgc	27540
agacaacgca ccgaccgtgc cttcatcaa ccccccttc gtctcttcag atggattcca	27600
agagaagccc ctgggggtgt tgtccctgcg actggccgac cccgtcacca ccaagaacgg	27660
ggaaatcacc ctcaagctgg gagaggggggt ggacctgat tcctcgggaa aactcatctc	27720
caacacggcc accaaggccg ccgcccctct cagtttttcc aacaacacca tttcccttaa	27780
catggatcac cccttttaca ctaaagatgg aaaattatcc ttacaagttt ctccaccatt	27840
aaatatactg agaacaagca ttctaaacac actagcttta ggttttggat caggtttagg	27900
actccgtggc tctgccttgg cagtacagtt agtctctcca cttacatttg atactgatgg	27960

aaacataaag	cttaccttag	acagaggttt	gcatgttaca	acaggagatg	caattgaaag	28020
caacataagc	tgggctaaag	gtttaaaatt	tgaagatgga	gccatagcaa	ccaacattgg	28080
aatgggtta	gagtttgaa	gcagtagtac	agaacaggt	gttgatgatg	cttacccaat	28140
ccaagttaaa	cttgatctg	gccttagctt	tgacagtaca	ggagccataa	tggttggtaa	28200
caaagaagac	gataaactca	ctttgtggac	aacacctgat	ccatcaccaa	actgtcaaat	28260
actcgagaa	aatgatgcaa	aactaacact	ttgcttgact	aatgtggta	gtcaaatact	28320
ggccactgtg	tcagtcttag	ttgtaggaag	tgaaaccta	aacccatta	ctggcaccgt	28380
aagcagtgt	caggtgtttc	tacgttttga	tgcaaacggt	gttcttttaa	cagaacattc	28440
tactataaaa	aaatactggg	ggtataggca	gggagatagc	atagatggca	ctccatatac	28500
caatgctgta	ggattcatgc	ccaatttaaa	agcttatcca	aagtcacaaa	gttctactac	28560
taaaaataat	atagtagggc	aagtatacat	gaatggagat	gtttcaaaac	ctatgcttct	28620
cactataacc	ctcaatggta	ctgatgacag	caacagtaca	tattcaatgt	cattttcata	28680
cacctggact	aatggaagct	atgttggagc	aacatttggg	gctaactctt	ataccttctc	28740
atacatcgcc	caagaatgaa	cactgtatcc	caccctgcat	gccaaccctt	cccacccac	28800
tctgtggaac	aaactctgaa	acacaaaata	aaataaagtt	caagtgtttt	attgattcaa	28860
cagttttaca	ggattcgagc	agttatTTTT	cctccaccct	cccaggacat	ggaatacacc	28920
accctctccc	ccgcacagc	cttgaacatc	tgaatgccat	tggtgatgga	catgcttttg	28980
gtctccacgt	tccacacagt	ttcagagcga	gccagtctcg	ggtcggtcag	ggagatgaaa	29040
ccctccgggc	actcccgcat	ctgcacctca	cagctcaaca	gctgaggatt	gtcctcggtg	29100
gtcgggatca	cggttatctg	gaagaagcag	aagagcggcg	gtgggaatca	tagtccgcga	29160
acgggatcgg	ccggtggtgt	cgcatcaggc	ccgcagcag	tcgctgccgc	cgccgctccg	29220
tcaagctgct	gctcaggggg	tccgggtcca	gggactccct	cagcatgatg	cccacggccc	29280
tcagcatcag	tcgtctggtg	cggcgggcgc	agcagcgcat	gcggatctcg	ctcaggtcgc	29340
tgcagtacgt	gcaacacaga	accaccaggt	tgttcaacag	tccatagttc	aacacgctcc	29400
agccgaaact	catcgcggga	aggatgctac	ccacgtggcc	gtcgtaccag	atcctcaggt	29460
aatcaagtgt	gtgccccctc	cagaacacgc	tgcccacgta	catgatctcc	ttgggcatgt	29520
ggcggttcac	cacctcccg	taccacatca	ccctctgggt	gaacatgcag	ccccggatga	29580

tcctgcggaa ccacagggcc agcaccgccc cgcccgccat gcagcgaaga gacccccgggt	29640
cccggcaatg gcaatggagg acccaccgct cgtacccgtg gatcatctgg gagctgaaca	29700
agtctatgtt ggcacagcac aggcataatgc tcatgcatct cttcagcact ctcaactcct	29760
cgggggtcaa aaccatatcc cagggcacgg ggaactcttg caggacagcg aacccccgag	29820
aacagggcaa tcctcgaca gaacttacat tgtgcatgga cagggtatcg caatcaggca	29880
gcaccgggtg atcctccacc agagaagcgc gggctctcgt ctcctcacag cgtggtaagg	29940
gggccggccg atacgggtga tggcgggacg cggctgatcg tggttcgcgac cgtgtcatga	30000
tgcatgtgct ttcggacatt ttcgtacttg ctgtagcaga acctgggtccg ggcgctgcac	30060
accgatcgcc ggcggcggtc tcggcgcttg gaacgctcgg tggtgaaatt gtaaaacagc	30120
cactctctca gaccgtgcag cagatctagg gcctcaggag tgatgaagat cccatcatgc	30180
ctgatggctc tgatcacatc gaccaccgtg gaatgggcca gaccagcca gatgatgcaa	30240
ttttgttggg tttcggtgac ggcgggggag ggaagaacag gaagaacat gattaacttt	30300
taatccaaac ggtctcggag tacttcaaaa tgaagatcgc ggagatggca cctctcgccc	30360
ccgctgtgtt ggtggaaaat aacagccagg tcaaaggtga tacggttctc gagatgttcc	30420
acggtggctt ccagcaaagc ctccacgcgc acatccagaa acaagacaat agcgaaagcg	30480
ggagggttct ctaattcctc aatcatcatg ttacactcct gcaccatccc cagataattt	30540
tcatttttcc agccttgaat gattcgaact agttcgtgag gtaaatacaa gccagccatg	30600
ataaagagct cgcgcagagc gccctccacc ggcattctta agcacaccct cataattcca	30660
agatattctg ctcctggttc acctgcagca gattgacaag cggaatatca aaatctctgc	30720
cgcgatccct gagctcctcc ctcagcaata actgtaagta ctctttcata tcctctccga	30780
aattttttagc cataggacca ccaggaataa gattagggca agccacagta cagataaacc	30840
gaagtcctcc ccagtgcgca ttgccaaatg caagactgct ataagcatgc tggctagacc	30900
cggatgatatc ttccagataa ctggacagaa aatcgcccag gcaattttta agaaaatcaa	30960
caaaagaaaa atcctccagg tggacgttta gagcctcggg aacaacgatg aagtaaattgc	31020
aagcggtgcg ttccagcatg gttagtttagc tgatctgtag aaaaaacaaa aatgaacatt	31080
aaaccatgct agcctggcga acaggtgggt aaatcgttct ctccagcacc aggcaggcca	31140
cgggggtctcc ggcgcgaccc tcgtaaaaaat tgctgctatg attgaaaacc atcacagaga	31200

gacgttccccg gtggccggcg tgaatgattc gacaagatga atacaccccc ggaacattgg	31260
cgtccgcgag tgaaaaaaag cgcccagga agcaataagg cactacaatg ctcagtctca	31320
agtccagcaa agcgatgcc a tgccgatgaa gcacaaaatt ctccaggtgcg tacaaaatgt	31380
aattactccc ctctgcaca ggcagcaaag ccccgatcc ctccaggtac acatacaaag	31440
cctcagcgtc catagcttac cgagcagcag cacacaacag gcgcaagagt cagagaaagg	31500
ctgagctcta acctgtccac ccgctctctg ctcaatatat agcccagatc tacactgacg	31560
taaaggccaa agtctaaaaa taccgcgcaa ataatcacac acgcccagca cacgcccaga	31620
aaccggtgac acactcaaaa aaatacgcgc acttcctcaa acgccccaaa ctgccgtcat	31680
ttccgggttc ccacgctacg tcatcaaaac acgactttca aattccgtcg accgttaaaa	31740
acgtcacccg ccccgcccct aacggtcgcc cgtctctcag ccaatcagcg ccccgcatcc	31800
ccaaattcaa acacctcatt tgcatattaa cgcgcacaaa aagtttgagg tatattattg	31860
atgatgg	31867

<210> 12

<211> 32788

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic
polynucleotide

<400> 12

ccatcttcaa taatatacct caaacttttt gtgcgcgtta atatgcaaat gaggcgtttg	60
aatttgggga ggaagggcgg tgattggctg agggatgagc gaccgttagg ggcggggcga	120
gtgacgtttt gatgacgtgg ttgcgaggag gagccagttt gcaagttctc gtgggaaaag	180
tgacgtcaaa cgagggtgtg tttgaacacg gaaataactca attttccgc gctctctgac	240
aggaaatgag gtgtttctgg gcggatgcaa gtgaaaacgg gccattttcg cgcgaaaact	300
gaatgaggaa gtgaaaatct gagtaatttc gcgtttatgg caggaggagg tatttgccga	360
gggccgagta gactttgacc gattacgtgg gggtttcgat taccgtgttt ttcacctaaa	420
tttccgcgta cgggtgtcaaa gtccggtgtt tttacgtagg tgtcagctga tcgccagggt	480
atttaaacct gcgctctcca gtcaagaggc cactcttgag tgccagcgag aagagttttc	540
tcctccgcgc cgcgagtcag atctacactt tgaaagtagg gataacaggg taatgacatt	600

gattattgac tagttgttaa tagtaatcaa ttacggggtc attagttcat agcccatata	660
tggagttccg cgttacataa cttacggtaa atggcccgcc tggctgaccg cccaacgacc	720
ccgcccatt gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc	780
attgacgtca atgggtggag tattttacggt aaactgcccc cttggcagta catcaagtgt	840
atcatatgcc aagtccgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt	900
atgcccagta catgacctta cgggactttc ctacttggca gtacatctac gtattagtca	960
tcgctattac catggtgatg cggttttggc agtacaccaa tgggcgtgga tagcggtttg	1020
actcacgggg atttccaagt ctccacccca ttgacgtcaa tgggagtttg ttttggcacc	1080
aaaatcaacg ggactttcca aaatgtcgta ataacccgc cccgttgacg caaatgggcg	1140
gtaggcgtgt acgggtggag gtctatataa gcagagctcg tttagtgaac cgtcagatcg	1200
cctggaacgc catccacgct gttttgacct ccatagaaga cagcgatcgc gccaccatgg	1260
ccgggatgtt ccaggcactg tccgaaggct gcacacccta tgatattaac cagatgctga	1320
atgtcctggg agaccaccag gtctctggcc tggagcagct ggagagcatc atcaacttcg	1380
agaagctgac cgagtggaca agctccaatg tgatgcctat cctgtcccca ctgaccaagg	1440
gcatcctggg cttcgtgttt accctgacag tgccttctga gcggggcctg tcttgcac	1500
gcgaggcaga cgcaaccaca ccagagtccg ccaatctggg cgaggagatc ctgtctcagc	1560
tgtacctgtg gccccgggtg acatatcact ccccttctta cgctatcac cagttcgagc	1620
ggagagccaa gtacaagaga cacttcccag gctttggcca gtctctgctg ttcggctacc	1680
ccgtgtacgt gttcggcgat tgcgtgcagg gcgactggga tgccatccgg tttagatact	1740
gcgcaccacc tggatatgca ctgctgaggt gtaacgacac caattattcc gccctgctgg	1800
cagtgggagc cctggagggc cctcgcaatc aggattggct gggcgtgcca aggcagctgg	1860
tgacacgcat gcaggccatc cagaacgcag gcctgtgcac cctggtggca atgctggagg	1920
agacaatctt ctggctgcag gcctttctga tggccctgac cgacagcggc cccaagacaa	1980
acatcatcgt ggattcccag tacgtgatgg gcatctccaa gccttctttc caggagtttg	2040
tggactggga gaacgtgagc ccagagctga attccaccga tcagccattc tggcaggcag	2100
gaatcctggc aaggaacctg gtgcctatgg tggccacagt gcagggccag aatctgaagt	2160
accagggcca gagcctggtc atcagcgcct ccatcatcgt gtttaacctg ctggagctgg	2220

agggcgacta	tcgggacgat	ggcaacgtgt	gggtgcacac	cccactgagc	cccagaacac	2280
tgaacgcctg	ggtgaaggcc	gtggaggaga	agaagggcac	cccagtgcac	ctggagctgg	2340
cctccatgac	caatatggag	ctgatgtcta	gcatcgtgca	ccagcaggtg	aggacatacg	2400
gacccgtgtt	catgtgcctg	ggaggcctgc	tgaccatggg	ggcaggagcc	gtgtggctga	2460
cagtgcgggt	gctggagctg	ttcagagccg	cccagctggc	caacgatgtg	gtgctgcaga	2520
tcatggagct	gtgcggagca	gcctttcgcc	aggtgtgcca	caccacagtg	ccatggccca	2580
atgcctccct	gacccccaa	tggaacaatg	agacaacaca	gcctcagatc	gccaactgta	2640
gcgtgtacga	cttcttcgtg	tggctgcact	actatagcgt	gagggaatac	ctgtggcccc	2700
gcgtgacata	ccacatgaat	aagtacgcct	atcacatgct	ggagaggcgc	gccaagtata	2760
agagaggccc	tggcccaggc	gcaaagtttg	tggcagcatg	gaccctgaag	gccgccgccg	2820
gccccggccc	cggccagtat	atcaaggcta	acagtaagtt	cattggaatc	acagagctgg	2880
gacccggacc	tggataatga	gtttaaactc	ccatttaaata	gtgagggtta	atgcttcgag	2940
cagacatgat	aagatacatt	gatgagtttg	gacaaaccac	aactagaatg	cagtgaaaaa	3000
aatgctttat	ttgtgaaatt	tgtgatgcta	ttgctttatt	tgttaaccatt	ataagctgca	3060
ataaacaagt	taacaacaac	aattgcattc	attttatgtt	tcaggttcag	ggggagatgt	3120
gggagggtttt	ttaaagcaag	taaaacctct	acaaatgtgg	taaaataact	ataacgggtcc	3180
taaggtagcg	agtgagtagt	gttctggggc	gggggaggac	ctgcatgagg	gccagaataa	3240
ctgaaatctg	tgcttttctg	tgtgttgagc	cagcatgagc	ggaagcggct	cctttgaggg	3300
aggggtattc	agcccttatc	tgacggggcg	tctcccctcc	tgggcgggag	tgcgtcagaa	3360
tgtgatggga	tccacggtgg	acggccggcc	cgtgcagccc	gcgaactctt	caaccctgac	3420
ctatgcaacc	ctgagctctt	cgtcgttgga	cgcagctgcc	gccgcagctg	ctgcatctgc	3480
cgccagcgcc	gtgcgcggaa	tggccatggg	cgccggctac	tacggcactc	tggtggccaa	3540
ctcgagttcc	accaataatc	ccgccagcct	gaacgaggag	aagctgttgc	tgctgatggc	3600
ccagctcgag	gccttgaccc	agcgcctggg	cgagctgacc	cagcaggtgg	ctcagctgca	3660
ggagcagacg	cgggccgcgg	ttgccacggt	gaaatccaaa	taaaaaatga	atcaataaat	3720
aaacggagac	ggttggtgat	tttaacacag	agtctgaatc	tttatttgat	ttttcgcgcg	3780
cggtaggccc	tggaccaccg	gtctcgatca	ttgagcacc	ggtggatctt	ttccaggacc	3840

cggtagaggt	gggcttggat	gttgaggtac	atgggcatga	gcccgtcccg	ggggtggagg	3900
tagctccatt	gcagggcctc	gtgctcgggg	gtggtgttgt	aaatcaccca	gtcatagcag	3960
gggcgcaggg	catggtgttg	cacaatatct	ttgaggagga	gactgatggc	cacgggcagc	4020
cctttggtgt	aggtgtttac	aaatctgttg	agctgggagg	gatgcatgcg	gggggagatg	4080
aggtgcatct	tggcctggat	cttgagattg	gcgatgttac	cgcccagatc	ccgcctgggg	4140
ttcatgttgt	gcaggaccac	cagcacggtg	tatccggtgc	acttggggaa	tttatcatgc	4200
aacttggaag	ggaaggcgtg	aaagaatttg	gcgacgcctt	tgtgcccgcc	caggttttcc	4260
atgcactcat	ccatgatgat	ggcgatgggc	ccgtgggcgg	cggcctgggc	aaagacgttt	4320
cgggggtcgg	acacatcata	gttgtggtcc	tgggtgaggt	catcataggc	cattttaatg	4380
aatttggggc	ggagggtgcc	ggactggggg	acaaaggtag	cctcgatccc	gggggcgtag	4440
ttcccctcac	agatctgcat	ctcccaggct	ttgagctcgg	agggggggat	catgtccacc	4500
tgcggggcga	taaagaacac	ggtttccggg	gcgggggaga	tgagctgggc	cgaaagcaag	4560
ttccggagca	gctgggactt	gccgcagccg	gtggggccgt	agatgacccc	gatgaccggc	4620
tgcaggtggt	agttgaggga	gagacagctg	ccgtcctccc	ggaggagggg	ggccacctcg	4680
ttcatcatct	cgcgcacgtg	catgttctcg	cgaccaggtt	ccgccaggag	gcgctctccc	4740
cccagggata	ggagctcctg	gagcgaggcg	aagtttttca	gcggcttgag	tccgtcggcc	4800
atgggcattt	tggagagggt	ttgttgcaag	agttccaggc	ggtcccagag	ctcggatgatg	4860
tgctctacgg	catctcgatc	cagcagacct	cctcgtttcg	cgggttgagg	cggctgcggg	4920
agtagggcac	cagacgatgg	gcgtccagcg	cagccagggt	ccggtccttc	cagggtcgca	4980
gcgtccgcgt	cagggtggtc	tccgtcacgg	tgaaggggtg	cgcgccgggc	tgggcgcttg	5040
cgagggtgcg	cttcaggctc	atccggctgg	tcgaaaaccg	ctcccgatcg	gcgccctgcg	5100
cgtcggccag	gtagcaattg	accatgagtt	cgtagttgag	cgctcggcc	gcgtggcctt	5160
tggcgcgagg	cttacctttg	gaagtctgcc	cgcaggcggg	acagaggagg	gacttgaggg	5220
cgtagagctt	gggggcgagg	aagacggact	cgggggcgta	ggcgtccgcg	ccgcagtggg	5280
cgcagacggt	ctgcactcc	acgagccagg	tgaggtcggg	ctggtcgggg	tcaaaaacca	5340
gtttcccgcc	gttctttttg	atgcgtttct	tacctttggt	ctccatgagc	tcgtgtcccc	5400
gctgggtgac	aaagaggctg	tccgtgtccc	cgtagaccga	ctttatgggc	cggtcctcga	5460

gcggtgtgcc gcggtcctcc tcgtagagga accccgcccc ctccgagacg aaagcccggg	5520
tccaggccag cacgaaggag gccacgtggg acgggtagcg gtcgttgtcc accagcgggt	5580
ccaccttttc cagggtatgc aaacacatgt cccctcgtc cacatccagg aaggtgattg	5640
gcttgtaagt gtaggccacg tgaccggggg tcccgccgg gggggataa aagggtgcgg	5700
gtccctgctc gtcctcactg tcttcggat cgctgtccag gagcgccagc tgttggggta	5760
ggtattccct ctcgaaggcg ggcatgacct cggcactcag gttgtcagtt tctagaaacg	5820
aggaggattt gatattgacg gtgccggcgg agatgccttt caagagcccc tcgtccatct	5880
ggtcagaaaa gacgatcttt ttgttgtcga gcttggtggc gaaggagccg tagagggcgt	5940
tggagaggag cttggcgatg gagcgcatgg tctggttttt ttccttgtcg gcgcgctcct	6000
tggcggcgat gttgagctgc acgtactcgc gcgccacgca cttccattcg gggaagacgg	6060
tggtcagctc gtcgggcacg attctgacct gccagccccg attatgcagg gtgatgaggt	6120
ccacactggg gccacctcg ccgcgcagg gctcattagt ccagcagagg cgtccgccct	6180
tgcgcgagca gaaggggggc aggggggtcca gcatgacctc gtcggggggg tcggcatcga	6240
tggatgaagat gccgggcagg aggtcggggg caaagtagct gatggaagtg gccagatcgt	6300
ccagggcagc ttgccattcg cgcacggcca gcgcgctctc gtagggactg aggggcgtgc	6360
cccaggggcat gggatgggta agcgcggagg cgtacatgcc gcagatgtcg tagacgtaga	6420
ggggctcctc gaggatgccg atgtaggtgg ggtagcagcg ccccccgcgg atgctggcgc	6480
gcacgtagtc atacagctcg tgcgaggggg cgaggagccc cggggccagg ttggtgcgac	6540
tgggcttttc ggcgcggtag acgatctggc ggaaaatggc atgcgagttg gaggagatgg	6600
tgggcctttg gaagatgttg aagtgggcgt ggggcagtcc gaccgagtcg cggtatgaagt	6660
gggcgtagga gtcttgacg ttggcgacga gctcggcggg gactaggacg tccagagcgc	6720
agtagtcgag ggtctcctgg atgatgtcat acttgagctg tcccttttgt ttccacagct	6780
cgcggttag aaggaactct tcgcggtcct tccagtactc ttcgaggggg aaccctcct	6840
gatctgcacg gtaagagcct agcatgtaga actggttgac ggctttag ggcgagcagc	6900
ccttctccac ggggagggcg taggcctggg cggccttgcg cagggaggtg tgcgtgaggg	6960
cgaagtgct cctgacctg accttgagga actggtgctt gaagtcgata tcgtcgacg	7020
ccccctgctc ccagagctgg aagtccgtgc gcttcttgta ggcggggttg ggcaaagcga	7080

aagtaacatc gttgaagagg atcttgcccg cgcggggcat aaagttgcga gtgatgcgga	7140
aaggttgggg cacctcggcc cggttgttga tgacctgggc ggcgagcacg atctcgtcga	7200
agccgttgat gttgtggccc acgatgtaga gttccacgaa tcgcggacgg cccttgacgt	7260
ggggcagttt cttgagctcc tcgtaggtga gctcgtcggg gtcgctgagc ccgtgctgct	7320
cgagcgccca gtcggcgaga tgggggttgg cgcggaggaa ggaagtccag agatccacgg	7380
ccagggcggt ttgcagacgg tcccgggtact gacggaactg ctgcccgcg gccatTTTTT	7440
cgggggtgac gcagtagaag gtgcgggggt ccccgtgcca gcgatcccat ttgagctgga	7500
gggcgagatc gagggcgagc tcgacgagcc ggtcgtcccc ggagagtTtc atgaccagca	7560
tgaaggggac gagctgcttg ccgaaggacc ccatccaggt gtaggtttcc acatcgtagg	7620
tgaggaagag ctttccggtg cgaggatgcg agccgatggg gaagaactgg atctcctgcc	7680
accaattgga ggaatggctg ttgatgtgat ggaagtagaa atgccgacgg cgcgccgaac	7740
actcgtgctt gtgtttatac aagcggccac agtgctcgca acgctgcacg ggatgcacgt	7800
gctgcacgag ctgtacctga gttcctttga cgaggaattt cagtgggaag tggagtcgtg	7860
gcgcctgcat ctcgtgctgt actacgtcgt ggtggtcggc ctggccctct tctgcctcga	7920
tggtggtcat gctgacgagc ccgcgcggga ggcaggtcca gacctcggcg cgagcgggtc	7980
ggagagcgag gacgagggcg cgcaggccgg agctgtccag ggtcctgaga cgctgcggag	8040
tcaggtcagt gggcagcggc ggcgcgcggt tgacttgag gagtttttcc agggcgcgcg	8100
ggaggtccag atggtacttg atctccaccg cgccattggt ggcgacgtcg atggcttgca	8160
gggtcccgtg cccctgggggt gtgaccaccg tccccgttt cttcttgggc ggctggggcg	8220
acgggggchg tgcctcttcc atggttagaa gcggcggcga ggacgcgcgc cgggcggcag	8280
gggcggctcg gggcccgag gcaggggcgg caggggcacg tcggcgccgc gcgcgggtag	8340
gttctggtac tgcgccgga gaagactggc gtgagcgacg acgcgacggt tgacgtcctg	8400
gatctgacgc ctctgggtga aggccacggg acccgtagt ttgaacctga aagagagttc	8460
gacagaatca atctcggtat cgttgacggc ggcctgccgc aggatctctt gcacgtcgcc	8520
cgagttgtcc tggtaggcga tctcggtcat gaactgctcg atctcctcct cttgaaggtc	8580
tccgcggccg gcgcgctcca cgggtggccgc gaggtcgttg gagatgcggc ccatgagctg	8640
cgagaaggcg ttcatgccc cctcgtttca gacgcggctg tagaccacga cgccctcggg	8700

atcgcgggcg cgcatgacca cctgggcgag gttgagctcc acgtggcgcg tgaagaccgc	8760
gtagttgcag aggcgctggt agaggtagtt gagcgtggtg gcgatgtgct cggtagacgaa	8820
gaaatacatg atccagcggc ggagcggcat ctgctgacg tcgcccagcg cctccaaacg	8880
ttccatggcc tcgtaaaagt ccacggcgaa gttgaaaaac tgggagttgc gcgccgagac	8940
ggtcaactcc tcctccagaa gacggatgag ctggcgatg gtggcgcgca cctcgcgctc	9000
gaaggccccc gggagttcct ccacttcctc ttcttcctcc tccactaaca tctcttctac	9060
ttcttcctca ggcggcagtg gtggcggggg agggggcctg cgtcgccggc ggcgcacggg	9120
cagacggtcg atgaagcgct cgatggtctc gccgcgccgg cgtcgcatgg tctcggtgac	9180
ggcgcgcccc tcctcgcggg gccgcagcgt gaagacgccg ccgcgcatct ccaggtggcc	9240
gggggggtcc ccgttgggca gggagagggc gctgacgatg catcttatca attgccccgt	9300
agggactccg cgcaaggacc tgagcgtctc gagatccacg ggatctgaaa accgctgaac	9360
gaaggcttcg agccagtcgc agtcgcaagg taggctgagc acggtttctt ctggcgggtc	9420
atgttggttg ggagcggggc gggcgatgct gctggtgatg aagttgaaat aggcggttct	9480
gagacggcgg atggtggcga ggagcaccag gtctttgggc ccggcttgct ggatgcgcag	9540
acggtcggcc atgccccagg cgtggtcctg acacctggcc aggtccttgt agtagtcctg	9600
catgagccgc tccacgggca cctcctcctc gcccgcgcgg ccgtgcatgc gcgtgagccc	9660
gaagccgcgc tggggctgga cgagcgccag gtcggcgacg acgcgctcgg cgaggatggc	9720
ttgctggatc tgggtgaggg tggctctgga gtcacaaag tcgacgaagc ggtggtaggc	9780
tccggtgttg atggtgtagg agcagttggc catgacggac cagttgacgg tctggtggcc	9840
cggacgcacg agctcgtggt acttgaggcg cgagtaggcg cgcgtgtcga agatgtagtc	9900
gttgacagtg cgcaccaggt actggtagcc gatgaggaag tgcggcggcg gctggcggtg	9960
gagcggccat cgctcggtgg cgggggcgcc gggcgcgagg tcctcgagca tggtagcggtg	10020
gtagccgtag atgtacctgg acatccaggt gatgccggcg gcggtggtgg aggcgcgcgg	10080
gaactcgcgg acgcggttcc agatgttcg cagcggcagg aagtagttca tgggtgggcac	10140
ggtctggccc gtgaggcgcg cgcagtcgtg gatgctctat acgggcaaaa acgaaagcgg	10200
tcagcggctc gactccgtgg cctggaggct aagcgaacgg gttgggctgc gcgtgtaccc	10260
cggttcgaat ctgcaatcag gctggagccg cagctaacgt ggtattggca ctcccgtctc	10320

gaccaagcc tgcaccaacc ctccaggata cggaggcggg tcgttttgca actttttttt	10380
ggaggccgga tgagactagt aagcgcggaa agcggccgac cgcgatggct cgctgccgta	10440
gtctggagaa gaatcgccag ggttgcgttg cgggtgtgccc cggttcgagg ccggccggat	10500
tccgcggcta acgagggcgt ggctgccccg tcgtttccaa gaccccatag ccagccgact	10560
tctccagtta cggagcgagc ccctcttttg ttttgtttgt ttttgccaga tgcattcccgt	10620
actgcggcag atgcgcccc accaccctcc accgcaacaa cagccccctc cacagccggc	10680
gcttctgccc ccgccccagc agcaacttcc agccacgacc gccgcggccg ccgtgagcgg	10740
ggctggacag agttatgatc accagctggc cttggaagag ggcgaggggc tggcgcgcct	10800
gggggcgtcg tcgccggagc ggcacccgcg cgtgcagatg aaaagggacg ctgcgaggc	10860
ctacgtgccc aagcagaacc tgttcagaga caggagcggc gaggagcccg aggagatgcg	10920
cgcgccccgg ttccacgcgg ggcgggagct gcggcgcggc ctggaccgaa agagggtgct	10980
gagggacgag gatttcgagg cggacgagct gacggggatc agccccgcgc gcgcgcacgt	11040
ggccgcggcc aacctggtca cggcgtagca gcagaccgtg aaggaggaga gcaacttcca	11100
aaaatccttc aacaaccacg tgcgcaccct gatcgcgcg gcgagggtga ccctgggcct	11160
gatgcacctg tgggacctgc tggaggccat cgtgcagaac cccaccagca agccgctgac	11220
ggcgcagctg ttcctggtgg tgcagcatag tcgggacaac gaagcgttca gggaggcgct	11280
gctgaatatc accgagcccg agggccgctg gtcctggac ctggtgaaca ttctgcagag	11340
catcgtggtg caggagcgcg ggctgccgct gtccgagaag ctggcggcca tcaacttctc	11400
ggtgctgagt ttgggcaagt actacgctag gaagatctac aagacccgt acgtgcccac	11460
agacaaggag gtgaagatcg acgggtttta catgcgcatg accctgaaag tgctgaccct	11520
gagcgacgat ctgggggtgt accgcaacga caggatgcac cgtgcggtga gcgccagcag	11580
gcggcgcgag ctgagcgacc aggagctgat gcatagtctg cagcgggccc tgaccggggc	11640
cgggaccgag ggggagagct actttgacat gggcgcggac ctgcactggc agcccagccg	11700
ccgggccttg gaggcggcgg caggacccta cgtagaagag gtggacgatg aggtggacga	11760
ggagggcgag tacctggaag actgatggcg cgaccgtatt tttgctagat gcaacaacaa	11820
cagccacctc ctgatccgc gatgcgggcg gcgctgcaga gccagccgtc cggcattaac	11880
tcctcggacg attggacca ggccatgcaa cgcattcatg cgctgacgac ccgcaacccc	11940

gaagccttta	gacagcagcc	ccaggccaac	cggctctcgg	ccatcctgga	ggccgtggtg	12000
ccctcgcgct	ccaacccac	gcacgagaag	gtcctggcca	tcgtgaacgc	gctggtggag	12060
aacaaggcca	tccgcggcga	cgaggccggc	ctggtgtaca	acgcgctgct	ggagcgcgtg	12120
gcccgtaca	acagcaccaa	cgtgcagacc	aacctggacc	gcatggtgac	cgacgtgcgc	12180
gaggccgtgg	cccagcgcga	gcggttccac	cgcgagtcca	acctgggatc	catggtggcg	12240
ctgaacgcct	tcctcagcac	ccagcccgcc	aacgtgcccc	ggggccagga	ggactacacc	12300
aacttcatca	gcgccctgcg	cctgatggtg	accgaggtgc	cccagagcga	ggtgtaccag	12360
tccgggccgg	actacttctt	ccagaccagt	cgccagggct	tgcagaccgt	gaacctgagc	12420
caggctttca	agaacttgca	gggcctgtgg	ggcgtgcagg	ccccggtcgg	ggaccgcgcg	12480
acggtgtcga	gcctgctgac	gccgaactcg	cgctgtctgc	tgctgtctggt	ggcccccttc	12540
acggacagcg	gcagcatcaa	ccgcaactcg	tacctgggct	acctgattaa	cctgtaccgc	12600
gaggccatcg	gccaggcgca	cgtggacgag	cagacctacc	aggagatcac	ccacgtgagc	12660
cgcgccctgg	gccaggacga	ccggggcaac	ctggaagcca	ccctgaactt	tttgctgacc	12720
aaccggtcgc	agaagatccc	gccccagtac	gcgctcagca	ccgaggagga	gcgcatcctg	12780
cgttacgtgc	agcagagcgt	gggcctgttc	ctgatgcagg	agggggccac	ccccagcgcc	12840
gcgctcgaca	tgaccgcgcg	caacatggag	cccagcatgt	acgccagcaa	ccgcccgttc	12900
atcaataaac	tgatggacta	cttgcatcgg	gcggccgcca	tgaactctga	ctatttcacc	12960
aacgccatcc	tgaatcccca	ctggctcccg	ccgccgggggt	tctacacggg	cgagtacgac	13020
atgcccgacc	ccaatgacgg	gttcctgttg	gacgatgtgg	acagcagcgt	gttctcccc	13080
cgaccgggtg	ctaacgagcg	ccccttgttg	aagaaggaag	gcagcgaccg	acgcccgtcc	13140
tcggcgctgt	ccggccgcga	gggtgctgcc	gcggcggtgc	ccgaggccgc	cagtcctttc	13200
ccgagcttgc	ccttctcgct	gaacagtatc	cgcagcagcg	agctgggcag	gatcacgcgc	13260
ccgcgcttgc	tgggcgaaga	ggagtacttg	aatgactcgc	tgttgagacc	cgagcgggag	13320
aagaacttcc	ccaataacgg	gatagaaagc	ctggtggaca	agatgagccg	ctggaagacg	13380
tatgcgcagg	agcacaggga	cgatccccgg	gcgtcgcagg	ggggccacgag	ccggggcgagc	13440
gccgcccgtg	aacgccggtg	gcacgacagg	cagcggggac	agatgtggga	cgatgaggac	13500
tccgccgacg	acagcagcgt	gttggacttg	ggtgggagtg	gtaaccggtt	cgctcacctg	13560

cgccccgta tcggg'gcgat gatgtaagag aaaccgaaaa taaatgatac tcaccaaggc	13620
catggcgacc agcgtgcgtt cgtttcttct ctgttgttgt tgtatctagt atgatgaggc	13680
gtgcgtaccc ggagggtcct cctccctcgt acgagagcgt gatgcagcag gcgatggcgg	13740
cggcggcgat gcagcccccg ctggaggctc cttacgtgcc cccgcggtac ctggcgccta	13800
cggagggggcg gaacagcatt cgttactcgg agctggcacc cttgtacgat accacccggt	13860
tgtacctggt ggacaacaag tcggcggaca tcgcctcgct gaactaccag aacgaccaca	13920
gcaacttcct gaccaccgtg gtgcagaaca atgacttcac ccccacggag gccagcacc	13980
agaccatcaa ctttgacgag cgctcgcggt ggggcggcca gctgaaaacc atcatgcaca	14040
ccaacatgcc caacgtgaac gagttcatgt acagcaacaa gttcaaggcg cgggtgatgg	14100
tctcccgcaa gacccccaat ggggtgacag tgacagagga ttatgatggt agtcaggatg	14160
agctgaagta tgaatgggtg gaatttgagc tgcccgaagg caacttctcg gtgaccatga	14220
ccatcgacct gatgaacaac gccatcatcg acaattactt ggcggtgggg cggcagaacg	14280
gggtgctgga gagcgacatc ggcgatgaagt tcgacactag gaacttcagg ctgggctggg	14340
accccgtagc cgagctggtc atgcccgggg tgtagaccaa cgaggctttc catcccgata	14400
ttgtcttgct gcccggctgc ggggtggact tcaccgagag ccgcctcagc aacctgctgg	14460
gcattcgcaa gaggcagccc ttccaggaag gcttcagat catgtacgag gatctggagg	14520
ggggcaacat cccgcgctc ctggatgtcg acgcctatga gaaaagcaag gaggatgcag	14580
cagctgaagc aactgcagcc gtagctaccg cctctaccga ggtcaggggc gataattttg	14640
caagcgccgc agcagtggca gcggccgagg cggctgaaac cgaaagtaag atagtcattc	14700
agccggtgga gaaggatagc aagaacagga gctacaacgt actaccggac aagataaaca	14760
ccgcctaccg cagctggtac ctagcctaca actatggcga ccccgagaag ggcgtgcgct	14820
cctggacgct gctcaccacc tcggacgtca cctgcggcgt ggagcaagtc tactggtcgc	14880
tgcccgcacat gatgcaagac ccggtcacct tccgctccac gcgtcaagtt agcaactacc	14940
cgggtggtggg cgccgagctc ctgcccgtct actccaagag cttcttcaac gagcaggccg	15000
tctactcgca gcagctgcgc gccttcacct cgcttacgca cgtcttcaac cgcttccccg	15060
agaaccagat cctcgtccgc ccgcccgcgc ccaccattac caccgtcagt gaaaacgttc	15120
ctgctctcac agatcacggg accctgccgc tgcgcagcag tatccgggga gtccagcgcg	15180

tgaccgttac	tgacgccaga	cgccgcacct	gcccctacgt	ctacaaggcc	ctgggcatag	15240
tcgcgccg	cgctcctctg	agccgcacct	tctaaatgtc	cattctcatc	tcgcccagta	15300
ataacaccgg	ttggggcctg	cgcgcgccca	gcaagatgta	cggaggcgct	cgccaacgct	15360
ccacgcaaca	ccccgtgcgc	gtgcgcgggc	acttccgcgc	tccctggggc	gccctcaagg	15420
gccgcgtgcg	gtcgcgcacc	accgtcgacg	acgtgatcga	ccagggtggg	gccgacgcgc	15480
gcaactacac	ccccgccgcc	gcgcccgtct	ccaccgtgga	cgccgtcatc	gacagcgtgg	15540
tggccgacgc	gcgccggtac	gcccgcgcca	agagccggcg	gcggcgcatc	gcccggcggc	15600
accggagcac	ccccgccatg	cgcgcgggcg	gagccttgct	gcgcagggcc	aggcgcacgg	15660
gacgcagggc	catgctcagg	gcggccagac	gcgcggcttc	aggcgccagc	gccggcagga	15720
cccggagacg	cgcgggccacg	gcggcgggcag	cggccatcgc	cagcatgtcc	cgcccgcggc	15780
gagggaaacgt	gtactgggtg	cgcgacgccg	ccaccggtgt	gcgcgtgccc	gtgcgcaccc	15840
gccccctcg	cacttgaaga	tgttcacttc	gcgatgttga	tgtgtcccag	cggcgaggag	15900
gatgtccaag	cgcaaattca	aggaagagat	gctccaggtc	atcgcgctg	agatctacgg	15960
ccctgcggtg	gtgaaggagg	aaagaaagcc	ccgaaaatc	aagcgggtca	aaaaggacaa	16020
aaaggaagaa	gaaagtgatg	tggacggatt	ggtggagttt	gtgcgcgagt	tcgcccccg	16080
gcggcgcgctg	cagtggcgcg	ggcggaaggt	gcaaccggtg	ctgagaccgc	gcaccaccgt	16140
ggtcttcacg	cccggcgagc	gctccggcac	cgcttccaag	cgctcctacg	acgaggtgta	16200
cggggatgat	gatattctgg	agcaggcggc	cgagcgctg	ggcgagtttg	cttacggcaa	16260
gcgcagccgt	tccgcaccga	aggaagaggc	ggtgtccatc	ccgctggacc	acggcaacc	16320
cacgccgagc	ctcaagccc	tgaccttgca	gcaggtgctg	ccgaccgcgg	cgccgcgccg	16380
ggggttcaag	cgcgagggcg	aggatctgta	ccccaccatg	cagctgatgg	tgcccaagcg	16440
ccagaagctg	gaagacgtgc	tggagaccat	gaaggtggac	ccggacgtgc	agcccgaggt	16500
caaggtgcgg	cccatcaagc	aggtggcccc	gggcctgggc	gtgcagaccg	tggacatcaa	16560
gattcccacg	gagcccatgg	aaacgcagac	cgagcccatg	atcaagccca	gcaccagcac	16620
catggaggtg	cagacggatc	cctggatgcc	atcggtcctt	agtcgaagac	cccggcgcaa	16680
gtacggcgcg	gccagcctgc	tgatgcccc	ctacgcgctg	catccttcca	tcatccccac	16740
gccgggctac	cgcggcacgc	gcttctaccg	cggtcatacc	agcagccgcc	gccgcaagac	16800

caccactcgc	cgccgccgtc	gccgcaccgc	cgctgcaacc	acccctgccg	ccctggtgcg	16860
gagagtgtac	cgccgcggcc	gcgcacctct	gaccctgccg	cgcgcgcgct	accacccgag	16920
catcgccatt	taaactttcg	cctgctttgc	agatcaatgg	ccctcacatg	ccgccttcgc	16980
gttcccatta	cgggctaccg	aggaagaaaa	ccgcgccgta	gaaggctggc	ggggaacggg	17040
atgcgtcgcc	accaccaccg	gcggcgggcg	gccatcagca	agcggttggg	gggaggcttc	17100
ctgcccgcgc	tgatcccat	catcgccgcg	gcgatcgggg	cgatccccgg	cattgcttcc	17160
gtggcggtgc	aggcctctca	gcgccactga	gacacacttg	gaaacatctt	gtaataaacc	17220
aatggactct	gacgctcctg	gtcctgtgat	gtgttttcgt	agacagatgg	aagacatcaa	17280
tttttcgtcc	ctggctccgc	gacacggcac	gcggccgttc	atgggcacct	ggagcgacat	17340
cggcaccagc	caactgaacg	ggggcgccct	caattggagc	agtctctgga	gcgggcttaa	17400
gaatttcggg	tccacgctta	aaacctatgg	cagcaaggcg	tggaacagca	ccacagggca	17460
ggcgctgagg	gataagctga	aagagcagaa	cttccagcag	aagggtggcg	atgggctcgc	17520
ctcgggcatc	aacgggggtg	tggacctggc	caaccaggcc	gtgcagcggc	agatcaacag	17580
ccgcctggac	ccggtgccgc	ccgccggctc	cgtggagatg	ccgcagggtg	aggaggagct	17640
gcctcccctg	gacaagcggg	gcgagaagcg	accccgcccc	gatgcggagg	agacgctgct	17700
gacgcacacg	gacgagccgc	ccccgtacga	ggaggcggtg	aaactgggtc	tgcccaccac	17760
gcggcccatc	gcgcccctgg	ccaccggggg	gctgaaaccc	gaaaagcccc	cgaccctgga	17820
cttgcctcct	ccccagcctt	cccgcccctc	tacagtggct	aagcccctgc	cgccggtggc	17880
cgtggcccgc	gcgcgaccgc	ggggcaccgc	ccgccctcat	gcgaactggc	agagcactct	17940
gaacagcatc	gtgggtctgg	gagtgcagag	tgtgaagcgc	cgccgctgct	attaaaccta	18000
ccgtagcgct	taacttgctt	gtctgtgtgt	gtatgtatta	tgtcgccgcc	gccgctgtcc	18060
accagaagga	ggagtgaaga	ggcgcgtcgc	cgagttgcaa	gatggccacc	ccatcgatgc	18120
tgccccagtg	ggcgtacatg	cacatcgccg	gacaggacgc	ttcggagtac	ctgagtccgg	18180
gtctggtgca	gtttgcccgc	gccacagaca	cctacttcag	tctggggaac	aagtttagga	18240
accccacggg	ggcgcccacg	cacgatgtga	ccaccgaccg	cagccagcgg	ctgacgctgc	18300
gcttcgtgcc	cgtggaccgc	gaggacaaca	cctactcgta	caaagtgcgc	tacacgctgg	18360
ccgtgggcga	caaccgcgtg	ctggacatgg	ccagcaccta	ctttgacatc	cgcggcgtgc	18420

tggatcgggg	ccctagcttc	aaaccctact	ccggcaccgc	ctacaacagt	ctggccccc	18480
agggagcacc	caacacttgt	cagtggacat	ataaagccga	tggtgaaaact	gccacagaaa	18540
aaacctatac	atatggaaat	gcacccgtgc	agggcattaa	catcacaaaa	gatggtattc	18600
aacttggaac	tgacaccgat	gatcagccaa	tctacgcaga	taaaacctat	cagcctgaac	18660
ctcaagtggg	tgatgctgaa	tggcatgaca	tacttggtac	tgatgaaaag	tatggaggca	18720
gagctcttaa	gcctgatacc	aaaatgaagc	cttgttatgg	ttcttttgcc	aagcctacta	18780
ataaagaagg	aggtcaggca	aatgtgaaaa	caggaacagg	cactactaaa	gaatatgaca	18840
tagacatggc	tttctttgac	aacagaagtg	cggctgctgc	tggcctagct	ccagaaattg	18900
ttttgtatac	tgaaaatgtg	gatttgga	ctccagatac	ccatattgta	tacaaagcag	18960
gcacagatga	cagcagctct	tctattaatt	tgggtcagca	agccatgccc	aacagaccta	19020
actacattgg	tttcagagac	aactttatcg	ggctcatgta	ctacaacagc	actggcaata	19080
tgggggtgct	ggccggtcag	gcttctcagc	tgaatgctgt	ggttgacttg	caagacagaa	19140
acaccgagct	gtcctaccag	ctcttgcttg	actctctggg	tgacagaacc	cggtatttca	19200
gtatgtggaa	tcaggcgggtg	gacagctatg	atcctgatgt	gcgcattatt	gaaaatcatg	19260
gtgtggagga	tgaacttccc	aactattggt	tccctctgga	tgctgttggc	agaacagata	19320
cttatcaggg	aattaaggct	aatggaactg	atcaaaccac	atggaccaa	gatgacagtg	19380
tcaatgatgc	taatgagata	ggcaagggtg	atccattcgc	catggaaaatc	aacatccaag	19440
ccaacctgtg	gaggaacttc	ctctacgcca	acgtggccct	gtacctgccc	gactcttaca	19500
agtacacgcc	ggccaatggt	accctgccc	ccaacaccaa	cacctacgat	tacatgaacg	19560
gccgggtggt	ggcgccctcg	ctggtggact	cctacatcaa	catcggggcg	cgctggtcgc	19620
tggatcccat	ggacaacgtg	aacccttca	accaccaccg	caatgcgggg	ctgcgctacc	19680
gctccatgct	cctgggcaac	gggcgctacg	tgcccttcca	catccagggtg	ccccagaaat	19740
ttttcgccat	caagagcctc	ctgctcctgc	ccgggtccta	cacctacgag	tggaacttcc	19800
gcaaggacgt	caacatgatc	ctgcagagct	ccctcggcaa	cgacctgcgc	acggacgggg	19860
cctccatctc	cttcaccagc	atcaacctct	acgccacctt	cttccccatg	gcgcacaaca	19920
cggcctccac	gctcgaggcc	atgctgcgca	acgacaccaa	cgaccagtcc	ttcaacgact	19980
acctctcggc	ggccaacatg	ctctaccca	tcccgccaa	cgccaccaac	gtgcccctct	20040

ccatccccctc	gcgcaactgg	gccgccttcc	gcggctggtc	cttcacgcgt	ctcaagacca	20100
aggagacgcc	ctcgctgggc	tccgggttcg	accctactt	cgtctactcg	ggctccatcc	20160
cctacctcga	cggcaccttc	tacctcaacc	acaccttcaa	gaaggctctcc	atcaccttcg	20220
actcctccgt	cagctggccc	ggcaacgacc	ggctcctgac	gccaacgag	ttcgaaatca	20280
agcgcaccgt	cgacggcgag	ggctacaacg	tggcccagtg	caacatgacc	aaggactggg	20340
tcctgggtcca	gatgctggcc	cactacaaca	tcggctacca	gggcttctac	gtgcccagg	20400
gctacaagga	ccgcatgtac	tccttcttcc	gcaacttcca	gcccattgagc	cgccagggtg	20460
tggacgaggt	caactacaag	gactaccagg	ccgtcaccct	ggcctaccag	cacaacaact	20520
cgggcttcgt	cggctacctc	gcgcccacca	tgcgccagg	ccagccctac	cccgccaact	20580
accctaccc	gctcatcggc	aagagcgccg	tcaccagcgt	caccagaaa	aagttcctct	20640
gcgacagggt	catgtggcgc	atcccccttct	ccagcaactt	catgtccatg	ggcgcgctca	20700
ccgacctcgg	ccagaacatg	ctctatgcc	actccgcca	cgcgctagac	atgaatttcg	20760
aagtcgacc	catggatgag	tccacccttc	tctatgttgt	cttcgaagtc	ttcgacgtcg	20820
tccgagtgc	ccagccccac	cgcggcgtca	tcgaggccgt	ctacctgcgc	accccccttct	20880
cggccggtaa	cgccaccacc	taagctcttg	cttcttgcaa	gccatggccg	cgggctccgg	20940
cgagcaggag	ctcagggcc	tcattccgca	cctgggctgc	gggccctact	tcctgggcac	21000
cttcgataag	cgcttcccgg	gattcatggc	cccgcacaa	ctggcctgcg	ccatcgtcaa	21060
cacggccggc	cgcgagaccg	ggggcgagca	ctggctggcc	ttcgccctgga	acccgcgctc	21120
gaacacctgc	tacctcttcg	accccttcgg	gttctcggac	gagcgccctca	agcagatcta	21180
ccagttcgag	tacgagggcc	tgctgcgccg	cagcgccctg	gccaccgagg	accgctgcgt	21240
caccctggaa	aagtccaccc	agaccgtgca	gggtccgcgc	tcggccgcct	gcgggctctt	21300
ctgctgcatg	ttcctgcacg	ccttcgtgca	ctggcccgac	cgcccatgg	acaagaaccc	21360
caccatgaac	ttgctgacgg	gggtgcccc	cggcatgctc	cagtcgcccc	aggtggaacc	21420
caccctgcgc	cgcaaccagg	aggcgctcta	ccgcttcctc	aactcccact	ccgcctactt	21480
tcgctccac	cgcgcgcgca	tcgagaaggc	caccgccttc	gaccgcatga	atcaagacat	21540
gtaaaccgtg	tgtgtatgtt	aaatgtcttt	aataaacagc	actttcatgt	tacacatgca	21600
tctgagatga	tttatttaga	aatcgaaagg	gttctgccgg	gtctcgcat	ggcccgcg	21660

cagggacacg ttgcggaact ggtacttggc cagccacttg aactcgggga tcagcagttt	21720
gggcagcggg gtgtcgggga aggagtcggt ccacagcttc cggtcagtt gcagggcgcc	21780
cagcaggtcg ggcgcggaga tcttgaaatc gcagttggga cccgcgttct gcgcgcggga	21840
gttgcggtac acgggggttg agcactggaa caccatcagg gccgggtgct tcacgctcgc	21900
cagcaccgtc gcgtcggtag tgctctccac gtcgaggtcc tcggcgttgg ccatcccgaa	21960
gggggtcatc ttgcaggtct gccttcccat ggtgggcacg caccgggct tgtggttgca	22020
atgcagtgac agggggatca gcatcatctg ggcctggtcg gcgttcatcc ccgggtacat	22080
ggccttcatg aaagcctcca attgcctgaa cgcctgctgg gccttggctc cctcggtgaa	22140
gaagaccccg caggacttgc tagagaactg gttggtggcg caccggcgt cgtgcacgca	22200
gcagcgcgcg tcgttggttg ccagctgcac cacgtgcgc cccagcgggt tctgggtgat	22260
cttgccccg tcggggttct ccttcagcgc gcgtgccccg ttctcgtctg ccacatccat	22320
ctcgatcatg tgctccttct ggatcatggt ggtcccgtgc aggacccgca gcttgccctc	22380
ggcctcggtg caccgtgca gccacagcgc gcacccgggt cactcccagt tcttgtgggc	22440
gatctgggaa tgcgcgtgca cgaagccctg caggaagcgg cccatcatgg tggtcagggt	22500
cttggtgcta gtgaaggtca gcggaatgcc gcggtgctcc tcgttgatgt acaggtggca	22560
gatgcggcgg tacacctcgc cctgctcggg catcagctgg aagttggctt tcaggtcgggt	22620
ctccacgcgg tagcgggtcca tcagcatagt catgatttcc atacccttct cccaggccga	22680
gacgatgggc aggctcatag ggttcttcac catcatctta gcgctagcag ccgcggccag	22740
ggggtcgtc tcgtccaggg tctcaaagct ccgcttgccg tccttctcgg tgatccgcac	22800
cggggggtag ctgaagccca cggccgccag ctctctctcg gcctgtcttt cgtcctcgt	22860
gtcctggctg acgtcctgca ggaccacatg cttggtcttg cggggtttct tcttgggcgg	22920
cagcggcggc ggagatgttg gagatggcga gggggagcgc gagttctcgc tcaccactac	22980
tatctcttcc tcttcttggg ccgaggccac gcggcggtag gtatgtctct tcgggggcag	23040
aggcggaggc gacgggctct cgcgcgcgcg acttggcgga tggctggcag agccccttcc	23100
gcgttcgggg gtgcgctccc ggcggcgtc tgactgactt cctccgcggc cggccattgt	23160
gttctcctag ggaggaacaa caagcatgga gactcagcca tcgccaacct cgccatctgc	23220
ccccaccgcc gacgagaagc agcagcagca gaatgaaagc ttaaccgccc cgccgccag	23280

ccccgccacc	tccgacgcgg	ccgtcccaga	catgcaagag	atggaggaat	ccatcgagat	23340
tgacctgggc	tatgtgacgc	ccgcggagca	cgaggaggag	ctggcagtgc	gcttttcaca	23400
agaagagata	caccaagaac	agccagagca	ggaagcagag	aatgagcaga	gtcaggctgg	23460
gctcgagcat	gacggcgact	acctccacct	gagcgggggg	gaggacgcgc	tcatcaagca	23520
tctggcccgg	caggccacca	tcgtcaagga	tgcgctgctc	gaccgcaccg	aggtgcccct	23580
cagcgtggag	gagctcagcc	gcgccctacga	gttgaacctc	ttctcgccgc	gcgtgcccc	23640
caagcgccag	cccaatggca	cctgcgagcc	caaccgcgc	ctcaacttct	acccggtctt	23700
cgcggtgccc	gaggccctgg	ccacctacca	catctttttc	aagaaccaa	agatccccgt	23760
ctcctgccgc	gccaaccgca	ccgcgccga	cgcccttttc	aacctgggtc	ccggcgcccg	23820
cctacctgat	atgcctcct	tggaagaggt	tccaagatc	ttcgagggtc	tgggcagcga	23880
cgagactcgg	gccgcgaacg	ctctgcaagg	agaaggagga	gagcatgagc	accacagcgc	23940
cctggtcgag	ttggaaggcg	acaacgcgcg	gctggcggtg	ctcaaacgca	cggtcgagct	24000
gacctatttc	gcctaccggg	ctctgaacct	gcccccaaa	gtcatgagcg	cggtcatgga	24060
ccaggtgctc	atcaagcgcg	cgtcgcccct	ctccgaggac	gagggcagtc	aagactccga	24120
ggagggaag	cccgtggtca	gcgacgagca	gctggcccg	tggctgggtc	ctaatgctag	24180
tccccagagt	ttggaagagc	ggcgcaaact	catgatggcc	gtggtcctgg	tgaccgtgga	24240
gctggagtgc	ctgcgccgct	tcttcgccga	cgcgagagacc	ctgcgcaagg	tcgaggagaa	24300
cctgcactac	ctcttcaggc	acgggttcgt	gcgccaggcc	tgcaagatct	ccaacgtgga	24360
gctgaccaac	ctggtctcct	acatgggcat	cttgcacgag	aaccgcctgg	ggcagaacgt	24420
gctgcacacc	accctgcgcg	gggaggcccg	gcgcgactac	atccgcgact	gcgtctacct	24480
ctacctctgc	cacacctggc	agacgggcat	gggcgtgtgg	cagcagtgtc	tggaggagca	24540
gaacctgaaa	gagctctgca	agctcctgca	gaagaacctc	aagggtctgt	ggaccgggtt	24600
cgacgagcgc	accaccgcct	cggacctggc	cgacctcatt	ttccccgagc	gcctcaggct	24660
gacgctgcgc	aacggcctgc	ccgactttat	gagccaaagc	atgttgcaaa	actttcgctc	24720
tttcatcctc	gaacgctccg	gaatcctgcc	cgccacctgc	tccgcgctgc	cctcggactt	24780
cgtgccgctg	accttccgcg	agtgcccccc	gccgctgtgg	agccactgct	acctgctgcg	24840
cctggccaac	tacctggcct	accactcgga	cgtgatcgag	gacgtcagcg	gcgagggcct	24900

gctcgagtgc	cactgccgct	gcaacctctg	cacgccgcac	cgctccctgg	cctgcaaccc	24960
ccagctgctg	agcgagaccc	agatcatcgg	caccttcgag	ttgcaagggc	ccagcgaagg	25020
cgagggttca	gccgccaagg	gggggtctgaa	actcaccctg	gggctgtgga	cctcggccta	25080
cttgcgcaag	ttcgtgcccc	aggactacca	tcccttcgag	atcaggttct	acgaggacca	25140
atcccatccg	cccaaggccg	agctgtcggc	ctgcgtcatc	acccaggggg	cgatcctggc	25200
ccaattgcaa	gccatccaga	aatcccgcga	agaattcttg	ctgaaaaagg	gccgcgggggt	25260
ctacctcgac	ccccagaccg	gtgaggagct	caaccccggc	ttcccccagg	atgccccgag	25320
gaaacaagaa	gctgaaagtg	gagctgccgc	ccgtggagga	tttggaggaa	gactgggaga	25380
acagcagtca	ggcagaggag	gaggagatgg	aggaagactg	ggacagcact	caggcagagg	25440
aggacagcct	gcaagacagt	ctggaggaag	acgaggagga	ggcagaggag	gagggtggaag	25500
aagcagccgc	cgccagaccg	tcgtcctcgg	cggggggagaa	agcaagcagc	acggatacca	25560
tctccgctcc	gggtcgggggt	cccgctcgac	cacacagtag	atgggacgag	accggacgat	25620
tcccgaaccc	caccaccag	accggtaaaga	aggagcggca	gggatacaag	tcctggcggg	25680
ggcacaaaaa	cgccatcgtc	tcctgcttgc	aggcctgcgg	gggcaacatc	tccttcaccc	25740
ggcgctacct	gctctttcac	cgcgggggtga	actttccccg	caacatcttg	cattactacc	25800
gtcacctcca	cagcccctac	tactttcaag	aagaggcagc	agcagcagaa	aaagaccagc	25860
agaaaaccag	cagctagaaa	atccacagcg	gcggcagcag	gtggactgag	gatcgcggcg	25920
aacgagccgg	cgcaaaccg	ggagctgagg	aaccggatct	ttcccaccct	ctatgccatc	25980
ttccagcaga	gtcgggggca	ggagcaggaa	ctgaaagtca	agaaccgttc	tctgcgctcg	26040
ctcaccgca	gttgtctgta	tcacaagagc	gaagaccaac	ttcagcgcac	tctcaggagc	26100
gccgaggctc	tcttcaacaa	gtactgcgcg	ctcactctta	aagagtagcc	cgcgcccgcc	26160
cagtcgcaga	aaaaggcggg	aattacgtca	cctgtgccct	tcgccctagc	cgccctccacc	26220
catcatcatg	agcaaagaga	ttcccacgcc	ttacatgtgg	agctaccagc	cccagatggg	26280
cctggccgcc	ggtgccgccc	aggactactc	caccgcgatg	aattggctca	gcgccggggc	26340
cgcgatgatc	tcacgggtga	atgacatccg	cgcccaccga	aaccagatac	tcctagaaca	26400
gtcagcgctc	accgccacgc	cccgcaatca	cctcaatccg	cgtaattggc	ccgccgcctt	26460
ggtgtaccag	gaaattcccc	agcccacgac	cgtactactt	ccgcgagacg	cccaggccga	26520

agtccagctg actaactcag gtgtccagct ggcggggcggc gccaccctgt gtcgtcaccg	26580
ccccgctcag ggtataaagc ggctgggtgat ccggggcaga ggcacacagc tcaacgacga	26640
ggtaggtgagc tcttcgctgg gtctgcgacc tgacggagtc ttccaactcg ccggatcggg	26700
gagatcttcc ttcacgcctc gtcaggccgt cctgactttg gagagtctgt cctcgcagcc	26760
ccgctcgggt ggcacgcgca ctctccagtt cgtggaggag ttactccct cggctctactt	26820
caacccttc tccggctccc ccggccacta ccgggacgag ttcatcccga acttcgacgc	26880
catcagcgag tcgggtggacg gctacgattg aatgtcccat ggtggcgcag ctgacctagc	26940
tcggcttcga cacctggacc actgccgccg cttccgctgc ttcgctcggg atctcgccga	27000
gtttgcctac tttgagctgc ccgaggagca ccctcagggc ccggcccacg gtagtcggat	27060
cgtcgtcgaa gggggcctcg actccacct gcttcggatc ttcagccagc gtccgatcct	27120
ggtcgagcgc gagcaaggac agacccttct gactctgtac tgcatctgca accaccccgg	27180
cctgcatgaa agtctttgtt gtctgctgtg tactgagtat aataaaagct gagatcagcg	27240
actactccgg acttccgtgt gttcctgaat ccatcaacca gtctttgttc ttcaccggga	27300
acgagaccga gctccagctc cagtgtaacg ccacaagaa gtacctcacc tggctgttcc	27360
agggctcccc gatcgccgtt gtcaaccact gcgacaacga cggagtcctg ctgagcggcc	27420
ctgccaacct tactttttcc acccgagaa gcaagctcca gctcttccaa cccttcctcc	27480
ccgggaccta tcagtgcgtc tcgggaccct gccatcacac cttccacctg atcccgaata	27540
ccacagcgtc gctccccgt actaacaacc aaactaacct ccaccaacgc caccgtcgcg	27600
acggccacaa tacatgcccata ttagacta tgaggccgag ccacagcgac ccatgctccc	27660
cgctattagt tacttcaatc taaccggcgg agatgactga ccactggcc aacaacaacg	27720
tcaacgacct tctcctggac atggacggcc gcgcctcgga gcagcgactc gcccaacttc	27780
gcattcgcca gcagcaggag agagccgtca aggagctgca ggatgcggtg gccatccacc	27840
agtgcaagag aggcatcttc tgcctgggtga aacaggccaa gatctcctac gaggtcactc	27900
caaacgacca tcgcctctcc tacgagctcc tgcagcagcg ccagaagttc acctgcctgg	27960
tcggagtcaa ccccatcgtc atcaccagc agtctggcga taccaagggg tgcatccact	28020
gctcctgcga ctccccgac tgcgtccaca ctctgatcaa gaccctctgc ggctccgcg	28080
acctcctccc catgaactaa tcacccctt atccagtga ataaagatca tattgatgat	28140

gattttacag aaataaaaaa taatcatttg atttgaaata aagatacaat catattgatg	28200
atttgagttt aacaaaaaaa taaagaatca cttacttgaa atctgatacc aggtctctgt	28260
ccatgttttc tgccaacacc acttcactcc cctcttccca gctctggtac tgcaggcccc	28320
ggcgggctgc aaacttcctc cacacgctga aggggatgtc aaattcctcc tgtccctcaa	28380
tcttcatttt atcttctatc agatgtccaa aaagcgcgtc cgggtggatg atgacttcga	28440
ccccgtctac ccctacgatg cagacaacgc accgaccgtg cccttcacatca accccccctt	28500
cgtctcttca gatggattcc aagagaagcc cctgggggtg ttgtccctgc gactggccga	28560
ccccgtcacc accaagaacg gggaaatcac cctcaagctg ggagaggggg tggacctcga	28620
ttcctcgga aaactcatct ccaacacggc caccaaggcc gccgcccctc tcagtttttc	28680
caacaacacc atttccctta acatggatca ccccttttac actaaagatg gaaaattatc	28740
cttacaagtt tctccaccat taaatatact gagaacaagc attctaaaca cactagcttt	28800
aggttttgga tcaggtttag gactccgtgg ctctgccttg gcagtacagt tagtctctcc	28860
acttacattt gatactgatg gaaacataaa gcttacctta gacagaggtt tgcattgttac	28920
aacaggagat gcaattgaaa gcaacataag ctgggctaaa ggtttaaaat ttgaagatgg	28980
agccatagca accaacattg gaaatgggtt agagtttgga agcagtagta cagaaacagg	29040
tgttgatgat gcttacccaa tccaagttaa acttgatctt ggcccttagct ttgacagtac	29100
aggagccata atggctggta acaaagaaga cgataaactc actttgtgga caacacctga	29160
tccatcacca aactgtcaaa tactcgcaga aaatgatgca aaactaacac tttgcttgac	29220
taaatgtggg agtcaaatac tggccactgt gtcagtctta gttgtaggaa gtggaaacct	29280
aaacccatt actggcaccg taagcagtgc tcagggtgtt ctacgttttg atgcaaacgg	29340
tgttctttta acagaacatt ctacactaaa aaaatactgg gggatataggc agggagatag	29400
catagatggc actccatata ccaatgctgt aggattcatg cccaatttaa aagcttatcc	29460
aaagtcacaa agttctacta ctaaaaataa tatagtaggg caagtataca tgaatggaga	29520
tgtttcaaaa cctatgcttc tcactataac cctcaatggg actgatgaca gcaacagtac	29580
atattcaatg tcattttcat acacctggac taatggaagc tatgttggag caacatttgg	29640
ggctaactct tataccttct catacatcgc ccaagaatga acactgtatc ccaccctgca	29700
tgccaaccct tcccaccca ctctgtggaa caaactctga aacacaaaat aaaataaagt	29760

tcaagtgttt	tattgattca	acagttttac	aggattcgag	cagttatttt	tcctccaccc	29820
tcccaggaca	tggaatacac	caccctctcc	ccccgcacag	ccttgaacat	ctgaatgcca	29880
ttggtgatgg	acatgctttt	ggtctccacg	ttccacacag	tttcagagcg	agccagtctc	29940
gggtcggcca	gggagatgaa	accctccggg	cactcccgcg	tctgcacctc	acagctcaac	30000
agctgaggat	tgtcctcggg	ggtcgggatc	acggttatct	ggaagaagca	gaagagcggc	30060
ggtgggaatc	atagtccgcg	aacgggatcg	gccggtgggtg	tcgcatcagg	ccccgcagca	30120
gtcgctgccg	ccgccgctcc	gtcaagctgc	tgctcagggg	gtccgggtcc	agggactccc	30180
tcagcatgat	gcccacggcc	ctcagcatca	gtcgtctggg	gcggcggggc	cagcagcgca	30240
tgcggatctc	gctcaggtcg	ctgcagtagc	tgcaacacag	aaccaccagg	ttgttcaaca	30300
gtccatagtt	caacacgctc	cagccgaaac	tcctcgcggg	aaggatgcta	cccacgtggc	30360
cgtcgtacca	gatcctcagg	taaatcaagt	ggtgccccct	ccagaacacg	ctgcccacgt	30420
acatgatctc	cttgggcatg	tggcgggttc	ccacctcccg	gtaccacatc	accctctggg	30480
tgaacatgca	gccccggatg	atcctgcgga	accacagggc	cagcaccgcc	ccgcccgcga	30540
tgcagcgaag	agaccccggg	tcccggcaat	ggcaatggag	gacccaccgc	tcgtaccctg	30600
ggatcatctg	ggagctgaac	aagtctatgt	tggcacagca	caggcatatg	ctcatgcatc	30660
tcttcagcac	tctcaactcc	tcgggggttc	aaaccatata	ccagggcacg	gggaactctt	30720
gcaggacagc	gaaccccgcg	gaacagggca	atcctcgcac	agaacttaca	ttgtgcatgg	30780
acagggtatc	gcaatcaggc	agcaccgggt	gatcctccac	cagagaagcg	cgggtctcgg	30840
tctcctcaca	gcgtggtaag	ggggccggcc	gatacgggtg	atggcgggac	gcggctgatc	30900
gtgttcgcga	ccgtgtcatg	atgcagttgc	tttcggacat	tttcgtactt	gctgtagcag	30960
aacctgggtc	gggcgctgca	caccgatcgc	cggcggcggt	ctcggcgctt	ggaacgctcg	31020
gtgttgaaat	tgtaaaacag	ccactctctc	agaccgtgca	gcagatctag	ggcctcagga	31080
gtgatgaaga	tcccatcatg	cctgatggct	ctgatcacat	cgaccaccgt	ggaatgggcc	31140
agaccagcc	agatgatgca	attttgttgg	gtttcgggtga	cggcggggga	gggaagaaca	31200
ggaagaacca	tgattaactt	ttaatccaaa	cggctctcga	gtacttcaaa	atgaagatcg	31260
cggagatggc	acctctcgcc	cccgtgtgtg	tggtggaaaa	taacagccag	gtcaaagggtg	31320
atacggttct	cgagatgttc	cacggtggct	tccagcaaag	cctccacgcg	cacatccaga	31380

aacaagacaa tagcgaaagc gggagggttc tctaattcct caatcatcat gttacactcc	31440
tgcaccatcc ccagataatt ttcatttttc cagccttgaa tgattcgaac tagttcgtga	31500
ggtaaatcca agccagccat gataaagagc tcgcgagag cgccctccac cggcattctt	31560
aagcacaccc tcataattcc aagatattct gctcctgggt cacctgcagc agattgacaa	31620
gcggaatatc aaaatctctg ccgcatccc tgagctctc cctcagcaat aactgtaagt	31680
actctttcat atcctctccg aaatttttag ccataggacc accaggaata agattagggc	31740
aagccacagt acagataaac cgaagtcctc cccagtgcgc attgccaaat gcaagactgc	31800
tataagcatg ctggctagac ccggtgatat cttccagata actggacaga aaatcgccca	31860
ggcaattttt aagaaaatca acaaaagaaa aatcctccag gtggacgttt agagcctcgg	31920
gaacaacgat gaagtaaag caagcgggtgc gttccagcat ggtagtag ctgatctgta	31980
gaaaaaaca aaatgaacat taaaccatgc tagcctggcg aacagggtggg taaatcggtc	32040
tctccagcac caggcaggcc acgggggtctc cggcgcgacc ctcgtaaaaa ttgtcgctat	32100
gattgaaaac catcacagag agacgtttcc ggtggccggc gtgaatgatt cgacaagatg	32160
aatacacccc cggaacattg gcgtccgcga gtgaaaaaaa gcgcccaggg aagcaataag	32220
gcactacaat gctcagtctc aagtccagca aagcgatgcc atgcggatga agcacaaaat	32280
tctcagggtgc gtacaaaatg taattactcc cctcctgcac aggcagcaaa gccccgatc	32340
cctccaggta cacatacaaa gcctcagcgt ccatagctta ccgagcagca gcacacaaca	32400
ggcgcaagag tcagagaaag gctgagctct aacctgtcca cccgctctct gctcaatata	32460
tagcccagat ctacactgac gtaaaggcca aagtctaaaa atacccgcca aataatcaca	32520
cacgcccagc acacgcccag aaaccgggtga cacactcaaa aaaatacgcg cacttcctca	32580
aacgccc aaa actgccgtca tttccgggtt cccacgctac gtcataaaaa cacgactttc	32640
aaattccgtc gaccgttaaa aacgtcaccg gccccgcccc taacgggtcgc ccgtctctca	32700
gccaatcagc gccccgcac cccaaattca aacacctcat ttgcatatta acgcgacaaa	32760
aaagtttgag gtatattatt gatgatgg	32788

<210> 13

<211> 30684

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic
polynucleotide

<400> 13

ccatcttcaa taatatacct caaacttttt gtgcgcgtta atatgcaaat gaggcgtttg	60
aatttgggga ggaagggcgg tgattggtcg agggatgagc gaccgttagg ggcggggcga	120
gtgacgtttt gatgacgtgg ttgcgaggag gagccagttt gcaagttctc gtgggaaaag	180
tgacgtcaaa cgaggtgtgg tttgaacacg gaaataactca attttcccgc gctctctgac	240
aggaaatgag gtgtttctgg gcggatgcaa gtgaaaacgg gccattttcg cgcgaaaact	300
gaatgaggaa gtgaaaatct gagtaatttc gcgtttatgg cagggaggag tatttgccga	360
gggccgagta gactttgacc gattacgtgg gggtttcgat taccgtgttt ttcacctaaa	420
tttccgcgta cgggtgtcaaa gtccggtgtt tttacgtagg tgtcagctga tcgccagggt	480
atttaaacct gcgctctcca gtcaagaggc cactcttgag tgccagcgag aagagttttc	540
tcctccgcgc cgcgagtcag atctacactt tgaaagtagg gataacaggg taatgacatt	600
gattattgac tagttgttaa tagtaatcaa ttacggggtc attagttcat agcccatata	660
tggagttccg cgttacataa cttacggtaa atggcccgcc tggctgaccg cccaacgacc	720
cccgcccatt gacgtcaata atgacgtatg ttcccatagt aacgccaata gggacttttc	780
attgacgtca atgggtggag tatttacggg aaactgcca cttggcagta catcaagtgt	840
atcatatgcc aagtccgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt	900
atgccagta catgacctta cgggactttc ctacttggca gtacatctac gtattagtca	960
tcgctattac catggtgatg cggttttggc agtacaccaa tgggcgtgga tagcggtttg	1020
actcacgggg atttccaagt ctccaccca ttgacgtcaa tgggagtttg ttttggcacc	1080
aaaatcaacg ggactttcca aaatgtcgta ataacccgc cccgttgacg caaatgggcg	1140
gtaggcgtgt acgggtgggag gtctatataa gcagagctcg tttagtgaac cgtcagatcg	1200
cctggaacgc catccacgct gttttgacct ccatagaaga cagcgatcgc gccaccatgg	1260
tgagcaaggg cgaggagctg ttcaccgggg tggtgcccat cctggtcgag ctggacggcg	1320
acgtaaacgg ccacaagttc agcgtgtccg gcgagggcga gggcgatgcc acctacggca	1380
agctgaccct gaagttcatc tgcaccaccg gcaagctgcc cgtgccctgg cccaccctcg	1440

tgaccaccct gacctacggc gtgcagtgtc tcagccgcta ccccgaccac atgaagcagc	1500
acgacttctt caagtccgcc atgcccgaag gctacgtcca ggagcgcacc atcttcttca	1560
aggacgacgg caactacaag acccgcgccg aggtgaagtt cgagggcgac accctggtga	1620
accgcatcga gctgaagggc atcgacttca aggaggacgg caacatcctg gggcacaagc	1680
tggagtacaa ctacaacagc cacaacgtct atatcatggc cgacaagcag aagaacggca	1740
tcaaggtgaa cttcaagatc cgccacaaca tcgaggacgg cagcgtgcag ctcgccgacc	1800
actaccagca gaacaccccc atcggcgacg gccccgtgtc gctgccccgac aaccactacc	1860
tgagcaccca gtccgccctg agcaaagacc ccaacgagaa gcgcgatcac atggtcctgc	1920
tggagttcgt gaccgccgcc gggatcactc tcggcatgga cgagctttac aagtagtgag	1980
tttaaactcc catttaaatt tgagggttaa tgcttcgagc agacatgata agatacattg	2040
atgagtttgg acaaaccaca actagaatgc agtgaaaaaa atgctttatt tgtgaaattt	2100
gtgatgctat tgctttattt gtaaccatta taagctgcaa taaacaagtt aacaacaaca	2160
attgcattca ttttatgttt caggttcagg gggagatgtg ggaggttttt taaagcaagt	2220
aaaacctcta caaatgtggt aaaataacta taacggctct aaggtagcga gtgagtagtg	2280
ttctggggcg ggggaggacc tgcatgaggg ccagaataac tgaaatctgt gcttttctgt	2340
gtgttgagc agcatgagcg gaagcggctc ctttgaggga ggggtattca gcccttatct	2400
gacggggcgt ctcccctcct gggcgggagt gcgtcagaat gtgatgggat ccacggtgga	2460
cggccggccc gtgcagcccg cgaactcttc aaccctgacc tatgcaaccc tgagctcttc	2520
gtcgttgga gtagctgccg ccgcagctgc tgcatctgcc gccagcgccg tgcgcggaat	2580
ggccatgggc gccggctact acggcactct ggtggccaac tcgagttcca ccaataatcc	2640
cgccagcctg aacgaggaga agctgttgct gctgatggcc cagctcgagg ccttgaccca	2700
gcgcctgggc gagctgaccc agcaggtggc tcagctgcag gagcagacgc gggccgcggt	2760
tgccacggtg aaatccaaat aaaaaatgaa tcaataaata aacggagacg gttgttgatt	2820
ttaacacaga gtctgaatct ttatttgatt tttcgcgcg ggtaggccct ggaccaccgg	2880
tctcgatcat tgagcacccg gtggatcttt tccaggaccc ggtagagggt ggcttgatg	2940
ttgaggtaca tgggcatgag cccgtcccgg gggtaggagt agctccattg cagggcctcg	3000
tgctcggggg tggtgttgta aatcaccag tcatagcagg ggcgcagggc atggtgttgc	3060

acaatatctt tgaggaggag actgatggcc acgggcagcc ctttgggtgta ggtgtttaca	3120
aatctgttga gctgggaggg atgcatgcgg ggggagatga ggtgcatctt ggcctggatc	3180
ttgagattgg cgatgttacc gccagatcc cgcctgggggt tcatgttggtg caggaccacc	3240
agcacggtgt atccggtgca cttggggaat ttatcatgca acttggaagg gaaggcgtga	3300
aagaatttgg cgacgccttt gtgcccggcc aggtttttcca tgcactcatc catgatgatg	3360
gcgatgggcc cgtgggaggc ggcctgggca aagacgtttc gggggtcgga cacatcatag	3420
ttgtggtcct gggtagagtc atcataggcc attttaatga atttggggcg gagggtgccg	3480
gactggggga caaagggtacc ctcgatcccg ggggcgtagt tcccctcaca gatctgcatc	3540
tcccaggctt tgagctcggga gggggggatc atgtccacct gcggggcgat aaagaacacg	3600
gtttccgggg cgggggagat gagctgggcc gaaagcaagt tccggagcag ctgggacttg	3660
ccgcagccgg tggggccgta gatgaccccg atgaccggct gcaggtggta gttgaggag	3720
agacagctgc cgtcctcccg gaggaggggg gccacctcgt tcatcatctc gcgcacgtgc	3780
atgttctcgc gcaccagttc cgccaggagg cgctctcccc ccagggatag gagctcctgg	3840
agcgaggcga agtttttcag cggcttgagt ccgtcggcca tgggcatttt ggagagggtt	3900
tgttgcaaga gttccaggcg gtcccagagc tcggtgatgt gctctacggc atctcgatcc	3960
agcagacctc ctcgtttcgc gggttgggac ggctgcggga gtagggcacc agacgatggg	4020
cgtccagcgc agccagggtc cggtccttcc agggtcgcag cgtccgcgtc aggggtggtct	4080
ccgtcacggg gaaggggtgc gcgccgggct gggcgcttgc gaggggtgcgc ttcagggtca	4140
tccggctggg cgaaaaccgc tcccgatcgg cgccctgcgc gtcggccagg tagcaattga	4200
ccatgagttc gtagttgagc gcctcggccg cgtggccttt ggcgcgagc ttacctttgg	4260
aagtctgccc gcaggcggga cagaggaggg acttgagggc gtagagcttg ggggcgagga	4320
agacggactc gggggcgtag gcgtccgcgc gcagtgggc gcagacggtc tcgcactcca	4380
cgagccaggt gaggtcgggc tggtcgggggt caaaaaccag tttcccgcg ttccttttga	4440
tgcgtttctt acctttggtc tccatgagct cgtgtccccg ctgggtgaca aagaggctgt	4500
ccgtgtcccc gtagaccgac tttatgggcc ggtcctcgag cgggtgtgccg cggtcctcct	4560
cgtagaggaa ccccgcccac tccgagacga aagcccgggt ccaggccagc acgaaggagg	4620
ccacgtggga cgggtagcgg tcgttgtcca ccagcgggtc caccttttcc agggtatgca	4680

aacacatgtc	cccctcgtcc	acatccagga	aggtgattgg	cttgtaagtg	taggccacgt	4740
gaccgggggt	cccggccggg	ggggtataaa	agggtgcggg	tccctgctcg	tcctcactgt	4800
cttccggatc	gctgtccagg	agcgccagct	gttggggtag	gtattccctc	tcgaaggcgg	4860
gcatgacctc	ggcactcagg	ttgtcagttt	ctagaaacga	ggaggatttg	atattgacgg	4920
tgccggcgga	gatgcctttc	aagagcccct	cgtccatctg	gtcagaaaag	acgatctttt	4980
tgttgtcgag	cttggtggcg	aaggagccgt	agagggcgtt	ggagaggagc	ttggcgatgg	5040
agcgcatggg	ctggtttttt	tccttgtcgg	cgcgctcctt	ggcggcgatg	ttgagctgca	5100
cgtactcgcg	cgccacgcac	ttccattcgg	ggaagacggg	ggtcagctcg	tcgggcacga	5160
ttctgacctg	ccagccccga	ttatgcaggg	tgatgaggtc	cacactgggtg	gccacctcgc	5220
cgcgcagggg	ctcattagtc	cagcagaggc	gtccgccctt	gcgcgagcag	aagggggggca	5280
gggggtccag	catgacctcg	tcgggggggt	cggcatcgat	ggtgaagatg	ccgggcagga	5340
ggtcggggtc	aaagtagctg	atggaagtgg	ccagatcgtc	cagggcagct	tgccattcgc	5400
gcacggccag	cgcgcgctcg	tagggactga	ggggcggtgc	ccagggcagc	ggatgggtaa	5460
gcgcggaggc	gtacatgccg	cagatgtcgt	agacgtagag	gggctcctcg	aggatgccga	5520
tgtaggtggg	gtagcagcgc	ccccgcgga	tgctggcgcg	cacgtagtca	tacagctcgt	5580
gcgagggggc	gaggagcccc	ggggccaggt	tggtgcgact	gggcttttcg	gcgcggtaga	5640
cgatctggcg	gaaaatggca	tgcgagttgg	aggagatggg	gggccttttg	aagatgttga	5700
agtgggcgtg	gggcagtccg	accgagtcgc	ggatgaagtg	ggcgtaggag	tcttgacgct	5760
tggcgacgag	ctcggcggtg	actaggacgt	ccagagcgca	gtagtcgagg	gtctcctgga	5820
tgatgtcata	cttgagctgt	cccttttggt	tccacagctc	gcggttgaga	aggaactctt	5880
cgcggtcctt	ccagtactct	tcgaggggga	acccgtcctg	atctgcacgg	taagagccta	5940
gcatgtagaa	ctggttgacg	gccttgtagg	cgcagcagcc	cttctccacg	gggagggcgt	6000
aggcctgggc	ggccttgctg	agggaggtgt	gcgtgagggc	gaaagtgtcc	ctgaccatga	6060
ccttgaggaa	ctgggtgctt	aagtcgatat	cgtcgcagcc	cccctgctcc	cagagctgga	6120
agtccgtgcg	cttcttgtag	gcgggggttg	gcaaagcgaa	agtaacatcg	ttgaagagga	6180
tcttgcccgc	gcggggcata	aagttgcgag	tgatgcggaa	aggttggggc	acctggcccc	6240
ggttggtgat	gacctgggcg	gcgagcacga	tctcgtcgaa	gccggtgatg	ttgtggccca	6300

cgatgtagag ttccacgaat cgcggacggc ccttgacgtg gggcagtttc ttgagctcct	6360
cgtaggtgag ctctgcgggg tcgctgagcc cgtgctgctc gagcgcccag tcggcgagat	6420
gggggttggc gcggaggaag gaagtccaga gatccacggc cagggcgggtt tgcagacggt	6480
cccgggtactg acggaactgc tgcccgaagg ccattttttc gggggtgacg cagtagaagg	6540
tgcgggggtc cccgtgccag cgatcccatt tgagctggag ggcgagatcg agggcgagct	6600
cgacgagccg gtcgtccccg gagagtttca tgaccagcat gaaggggacg agctgcttgc	6660
cgaaggaccc catccaggtg taggtttcca catcgtaggt gaggaagagc ctttcgggtgc	6720
gaggatgcga gccgatgggg aagaactgga tctcctgcc ccaattggag gaatggctgt	6780
tgatgtgatg gaagtagaaa tgccgacggc gcgccgaaca ctcgtgcttg tgtttataca	6840
agcggccaca gtgctcgcaa cgctgcacgg gatgcacgtg ctgcacgagc tgtacctgag	6900
ttcctttgac gaggaatttc agtgggaagt ggagtcgtgg cgcctgcatc tcgtgctgta	6960
ctacgtcgtg gtggtcggcc tggccctctt ctgcctcgat ggtggtcatt ctgacgagcc	7020
cgcgcgggag gcaggtccag acctcggcgc gagcgggtcg gagagcgagg acgagggcgc	7080
gcaggccgga gctgtccagg gtcctgagac gctgcggagt caggtcagtg ggcagcggcg	7140
gcgcgcgggt gacttgcagg agtttttcca gggcgcgcg gaggtccaga tgggtactga	7200
tctccaccgc gccattggtg gcgacgtcga tggcttgcag ggtcccgtgc ccctgggggtg	7260
tgaccaccgt cccccgtttc ttcttgggcg gctggggcga cgggggcgggt gcctcttcca	7320
tggttagaag cggcggcgag gacgcgcgcc gggcggcagg ggcggctcgg ggcccggagg	7380
caggggcggc aggggcacgt cggcgcgcgc cgcgggtagg ttctgggtact gcgcccggag	7440
aagactggcg tgagcgacga cgcgacgggt gacgtcctgg atctgacgcc tctgggtgaa	7500
ggccacggga cccgtgagtt tgaacctgaa agagagttcg acagaatcaa tctcggtatc	7560
gttgacggcg gcctgccga ggatctcttg cacgtcgccc gagttgtcct ggtaggcgat	7620
ctcggtcattg aactgctcga tctcctctc ttgaaggctt ccgcggccgg cgcgctccac	7680
ggtggccgcg aggtcgttgg agatgcggcc catgagctgc gagaaggcgt tcatgcccgc	7740
ctcgttccag acgcggctgt agaccacgac gccctcggga tcgcgggcgc gcatgaccac	7800
ctgggcgagg ttgagctcca cgtggcgcgt gaagaccgcg tagttgcaga ggcgctggta	7860
gaggtagttg agcgtggtgg cgatgtgctc ggtgacgaag aaatacatga tccagcggcg	7920

gagcggcatc	tcgctgacgt	cgcccagcgc	ctccaaacgt	tccatggcct	cgtaaaagtc	7980
cacggcgaag	ttgaaaaact	gggagttgcg	cgccgagacg	gtcaactcct	cctccagaag	8040
acggatgagc	tcggcgatgg	tggcgcgcac	ctcgcgctcg	aaggcccccg	ggagttcctc	8100
cacttcctct	tcttcctcct	ccactaacat	ctctttctact	tcctcctcag	gcggcagtgg	8160
tggcggggga	gggggcctgc	gtcgccggcg	gcgcacgggc	agacggtcga	tgaagcgctc	8220
gatggtctcg	ccgcgccggc	gtcgcatggt	ctcggtgacg	gcgcgcccgt	cctcgcgggg	8280
ccgcagcgtg	aagacgccgc	cgcgcatctc	caggtggccg	gggggggtccc	cgttgggcag	8340
ggagagggcg	ctgacgatgc	atcttatcaa	ttgccccgta	gggactccgc	gcaaggacct	8400
gagcgtctcg	agatccacgg	gatctgaaaa	ccgctgaacg	aaggcttcga	gccagtcgca	8460
gtcgcaaggt	aggctgagca	cggtttcttc	tggcgggtca	tgttggttgg	gagcggggcg	8520
ggcgatgctg	ctggtgatga	agttgaaata	ggcggttctg	agacggcgga	tggtggcgag	8580
gagcaccagg	tctttgggcc	cggcttgctg	gatgcgcaga	cggtcggcca	tgccccaggc	8640
gtggtcctga	cacctggcca	ggtccttgta	gtagtctctg	atgagccgct	ccacgggcac	8700
ctcctcctcg	cccgcgcggc	cgtgcatgcg	cgtgagcccc	aagccgcgct	ggggctggac	8760
gagcgccagg	tcggcgacga	cgcgctcggc	gaggatggct	tgctggatct	gggtgagggt	8820
ggtctggaag	tcatcaaagt	cgacgaagcg	gtggtaggct	ccggtgttga	tggtgtagga	8880
gcagttggcc	atgacggacc	agttgacggg	ctggtggccc	ggacgcacga	gctcgtggta	8940
cttgaggcgc	gagtaggcgc	gcgtgtcgaa	gatgtagtcg	ttgcagggtg	gcaccaggta	9000
ctggtagccg	atgaggaagt	gcggcggcgg	ctggcggtag	agcggccatc	gctcggtggc	9060
gggggcgccc	ggcgcgaggt	cctcgagcat	ggtgcgggtg	tagccgtaga	tgtacctgga	9120
catccaggtg	atgccggcgg	cgggtggtgga	ggcgcgcggg	aactcgcgga	cgcggttcca	9180
gatgttgcg	agcggcagga	agtagttcat	ggtgggcacg	gtctggcccc	tgaggcgcgc	9240
gcagtcgtgg	atgctctata	cgggcaaaaa	cgaaagcggg	cagcggctcg	actccgtggc	9300
ctggaggcta	agcgaacggg	ttgggctgcg	cgtgtacccc	ggttcgaatc	tcgaatcagg	9360
ctggagccgc	agctaacgtg	gtattggcac	tcccgtctcg	acccaagcct	gcaccaaccc	9420
tccaggatac	ggaggcgggt	cgttttgcaa	cttttttttg	gaggccggat	gagactagta	9480
agcgcggaaa	gcggccgacc	gcgatggctc	gctgccgtag	tctggagaag	aatcgccagg	9540

gttgcgttgc ggtgtgcccc ggttcgaggc cggccggatt ccgcggctaa cgagggcgtg	9600
gctgccccgt cgtttccaag accccatagc cagccgactt ctccagttac ggagcgagcc	9660
cctcttttgt tttgtttgtt tttgccagat gcatcccgtc ctgcggcaga tgcgccccca	9720
ccaccctcca ccgcaacaac agccccctcc acagccggcg cttctgcccc cgccccagca	9780
gcaacttcca gccacgaccg ccgcggccgc cgtgagcggg gctggacaga gttatgatca	9840
ccagctggcc ttggaagagg gcgaggggct ggcgcgctg ggggcgtcgt cgccggagcg	9900
gcaccgcgc gtgcagatga aaagggacgc tcgcgaggcc tacgtgcccc agcagaacct	9960
gttcagagac aggagcggcg aggagcccga ggagatgcgc gcggccccgt tccacgcggg	10020
gcgggagctg cggcgcgggc tggaccgaaa gagggtgctg agggacgagg atttcgaggc	10080
ggacgagctg acgggggatca gccccgcgcg cgcgcacgtg gccgcggcca acctggtcac	10140
ggcgtacgag cagaccgtga aggaggagag caacttccaa aaatccttca acaaccacgt	10200
gcgcaccctg atcgcgcgcg aggaggtgac cctgggcctg atgcacctgt gggacctgct	10260
ggaggccatc gtgcagaacc ccaccagcaa gccgctgacg gcgcagctgt tcctggtggt	10320
gcagcatagt cgggacaacg aagcgttcag ggaggcgctg ctgaatatca ccgagcccga	10380
gggccgctgg ctcttgacc tgggtgaacat tctgcagagc atcgtggtgc aggagcgcgg	10440
gctgccgctg tccgagaagc tggcggccat caacttctcg gtgctgagtt tgggcaagta	10500
ctacgctagg aagatctaca agaccccgta cgtgcccata gacaaggagg tgaagatcga	10560
cgggttttac atgcgcatga ccctgaaagt gctgaccctg agcgacgatc tgggggtgta	10620
ccgcaacgac aggatgcacc gtgcggtgag cgccagcagg cggcgcgagc tgagcgacca	10680
ggagctgatg catagtctgc agcggggcct gaccggggcc gggaccgagg gggagagcta	10740
ctttgacatg ggcgcgacc tgcaactggca gcccagccgc cgggccttgg aggcggcggc	10800
aggaccctac gtagaagagg tggacgatga ggtggacgag gagggcgagt acctggaaga	10860
ctgatggcgc gaccgtatctt ttgctagatg caacaacaac agccacctcc tgatccccgcg	10920
atgcggggcg cgctgcagag ccagccgtcc ggcattaact cctcggacga ttggaccag	10980
gcatgcaac gcatcatggc gctgacgacc cgcaaccccg aagcctttag acagcagccc	11040
caggccaacc ggctctcggc catcctggag gccgtggtgc cctcgcgctc caacccacg	11100
cacgagaagg tcctggccat cgtgaacgcg ctggtggaga acaaggccat ccgcggcgac	11160

gaggccggcc	tggtgtacaa	cgcgctgctg	gagcgcgtgg	cccgctacaa	cagcaccaac	11220
gtgcagacca	acctggaccg	catggtgacc	gacgtgcgcg	aggccgtggc	ccagcgcgag	11280
cggttccacc	gcgagtccaa	cctgggatcc	atggtggcgc	tgaacgcctt	cctcagcacc	11340
cagcccgcca	acgtgccccg	gggccaggag	gactacacca	acttcatcag	cgccctgcgc	11400
ctgatggtga	ccgaggtgcc	ccagagcgag	gtgtaccagt	ccgggccgga	ctacttcttc	11460
cagaccagtc	gccagggctt	gcagaccgtg	aacctgagcc	aggctttcaa	gaacttgcag	11520
ggcctgtggg	gcgtgcaggc	cccggtcggg	gaccgcgcga	cggtgtcgag	cctgctgacg	11580
ccgaactcgc	gcctgctgct	gctgctggtg	gcccccttca	cggacagcgg	cagcatcaac	11640
cgcaactcgt	acctgggcta	cctgattaac	ctgtaccgcg	aggccatcgg	ccaggcgcac	11700
gtggacgagc	agacctacca	ggagatcacc	cacgtgagcc	gcgccctggg	ccaggacgac	11760
ccgggcaacc	tggaagccac	cctgaacttt	ttgctgacca	accggtcgca	gaagatcccg	11820
ccccagtacg	cgctcagcac	cgaggaggag	cgcacacctg	gttacgtgca	gcagagcgtg	11880
ggcctgttcc	tgatgcagga	gggggccacc	cccagcgccg	cgctcgacat	gaccgcgcgc	11940
aacatggagc	ccagcatgta	cgccagcaac	cgcccgttca	tcaataaaact	gatggactac	12000
ttgcatcggg	cggccgccat	gaactctgac	tatttcacca	acgccatcct	gaatccccac	12060
tggtctccgc	cgccgggggtt	ctacacgggc	gagtacgaca	tgcccgaccc	caatgacggg	12120
ttcctgtggg	acgatgtgga	cagcagcgtg	ttctcccccc	gaccgggtgc	taacgagcgc	12180
cccttgtgga	agaaggaagg	cagcgaccga	cgcccgtcct	cggcgctgtc	cggccgcgag	12240
ggtgctgccg	cggcgggtgcc	cgaggccgcc	agtccttttc	cgagcttgcc	cttctcgtg	12300
aacagtatcc	gcagcagcga	gctgggcagg	atcacgcgcc	cgcgcttgct	gggcgaagag	12360
gagtacttga	atgactcgct	gttgagaccc	gagcgggaga	agaacttccc	caataacggg	12420
atagaaagcc	tggtggacaa	gatgagccgc	tggaagacgt	atgcgcagga	gcacagggac	12480
gatccccggg	cgtcgcaggg	ggccacgagc	cggggcagcg	ccgcccgtaa	acgccggtgg	12540
cacgacaggc	agcggggaca	gatgtgggac	gatgaggact	ccgccgacga	cagcagcgtg	12600
ttggacttgg	gtgggagtgg	taaccctgtt	gctcacctgc	gccccgtat	cgggcgcatg	12660
atgtaagaga	aaccgaaaat	aaatgatact	caccaaggcc	atggcgacca	gcgtgcgttc	12720
gtttcttctc	tgttgttgtt	gtatctagta	tgatgaggcg	tgcgtacccg	gagggtcctc	12780

ctccctcgta cgagagcgtg atgcagcagg cgatggcggc ggcggcgatg cagccccgc	12840
tggaggctcc ttacgtgccc ccgcggtacc tggcgctac ggaggggcgg aacagcattc	12900
gttactcgga gctggcacc ttgtacgata ccacccggtt gtacctggtg gacaacaagt	12960
cggcggacat cgcctcgctg aactaccaga acgaccacag caacttcctg accaccgtgg	13020
tgcagaacaa tgacttcacc cccacggagg ccagcaccca gaccatcaac tttgacgagc	13080
gctcgcggtg gggcggccag ctgaaaacca tcatgcacac caacatgccc aacgtgaacg	13140
agttcatgta cagcaacaag ttcaaggcgc gggatgatggc ctcccgcaag accccaatg	13200
gggtgacagt gacagaggat tatgatggta gtcaggatga gctgaagtat gaatgggtgg	13260
aatttgagct gcccgaggc aacttctcgg tgaccatgac catcgacctg atgaacaacg	13320
ccatcatcga caattacttg gcggtggggc ggcagaacgg ggtgctggag agcgacatcg	13380
gcgtgaagtt cgacactagg aacttcaggc tgggctggga ccccgtagc gagctggtca	13440
tgcccggggt gtacaccaac gaggccttcc atcccgatat tgtcttgctg cccggctgcg	13500
gggtggactt caccgagagc cgcctcagca acctgctggg cattcgcaag aggcagccct	13560
tccaggaagg cttccagatc atgtacgagg atctggaggg gggcaacatc cccgcgctcc	13620
tggatgtcga cgcctatgag aaaagcaagg aggatgcagc agctgaagca actgcagccg	13680
tagctaccgc ctctaccgag gtcaggggcg ataattttgc aagcgccgca gcagtggcag	13740
cggccgaggc ggctgaaacc gaaagtaaga tagtcattca gccgggtggag aaggatagca	13800
agaacaggag ctacaacgta ctaccggaca agataaacac cgcctaccgc agctggtacc	13860
tagcctacaa ctatggcgac cccgagaagg gcgtgcgctc ctggacgctg ctcaccacct	13920
cggacgtcac ctgcggcgtg gagcaagtct actggtcgct gcccgacatg atgcaagacc	13980
cggtcacctt ccgctccacg cgtcaagtta gcaactaccc ggtggtgggc gccgagctcc	14040
tgcccgtcta ctccaagagc ttcttcaacg agcaggccgt ctactcgag cagctgcgcg	14100
ccttcacctc gcttacgcac gtcttcaacc gcttccccga gaaccagatc ctcgtccgcc	14160
cgcgcgccc caccattacc accgtcagtg aaaacgttcc tgctctcaca gatcacggga	14220
ccctgccgct gcgcagcagt atccggggag tccagcgctg gaccgttact gacgccagac	14280
gccgcacctg cccctacgtc tacaaggccc tgggcatagt cgcgccgcgc gtcctctcga	14340
gccgcacctt ctaaagtcc attctcatct cggccagtaa taacaccggt tggggcctgc	14400

gcgcgcccag	caagatgtac	ggaggcgctc	gccaacgctc	cacgcaacac	cccgtgcgcg	14460
tgcgcgggca	cttccgcgct	ccctggggcg	ccctcaaggg	ccgcgtgcgg	tcgcgcacca	14520
ccgtcgacga	cgtgatcgac	caggtggtgg	ccgacgcgcg	caactacacc	cccgccgccg	14580
cgtccgtctc	caccgtggac	gccgtcatcg	acagcgtggt	ggccgacgcg	cgccggtacg	14640
cccgcgccaa	gagccggcgg	cggcgcatcg	cccggcggca	ccggagcacc	cccgccatgc	14700
gcgcggcgcg	agccttgctg	cgcagggcca	ggcgcacggg	acgcagggcc	atgctcaggg	14760
cggccagacg	cgcggttca	ggcgccagcg	ccggcaggac	ccggagacgc	gcggccacgg	14820
cggcggcagc	ggccatcgcc	agcatgtccc	gcccgcggcg	agggaacgtg	tactgggtgc	14880
gcgacgccgc	caccggtgtg	cgcgtgcccg	tgcgcacccg	ccccctcgc	acttgaagat	14940
gttcacttcg	cgatgttgat	gtgtcccagc	ggcgaggagg	atgtccaagc	gcaaattcaa	15000
ggaagagatg	ctccaggtca	tcgcgcctga	gatctacggc	cctgcggtgg	tgaaggagga	15060
aagaaagccc	cgcaaaatca	agcgggtcaa	aaaggacaaa	aaggaagaag	aaagtgatgt	15120
ggacggattg	gtggagtttg	tgcgcgagtt	cggcgcggtg	agtggcgcgg		15180
gcggaaggtg	caaccggtgc	tgagaccg	caccaccgtg	gtcttcacgc	ccggcgagcg	15240
ctccggcacc	gcttccaagc	gctcctacga	cgaggtgtac	ggggatgatg	atattctgga	15300
gcaggcggcc	gagcgctgg	gcgagtttgc	ttacggcaag	cgcagccgtt	ccgcaccgaa	15360
ggaagaggcg	gtgtccatcc	cgctggacca	cggcaacccc	acgccgagcc	tcaagcccgt	15420
gaccttgacg	caggtgctgc	cgaccgcggc	gccgcgccgg	gggttcaagc	gcgagggcga	15480
ggatctgtac	cccacatgc	agctgatggt	gcccgaagcg	cagaagctgg	aagacgtgct	15540
ggagaccatg	aaggtggacc	cggacgtgca	gcccgaagtc	aaggtgcggc	ccatcaagca	15600
ggtggccccg	ggcctggg	tgagaccgt	ggacatcaag	attcccacgg	agcccatgga	15660
aacgcagacc	gagcccatga	tcaagcccag	caccagcacc	atggaggtgc	agacggatcc	15720
ctggatgcca	tcggctccta	gtcgaagacc	ccggcgcaag	tacggcgcg	ccagcctgct	15780
gatgcccac	tacgcgtgc	atccttccat	catccccacg	ccgggctacc	gcggcacgcg	15840
cttctaccgc	ggtcatacca	gcagccgccg	ccgcaagacc	accactcgcc	gccgccgtcg	15900
ccgcaccgcc	gctgcaacca	cccctgccgc	cctggtgcgg	agagtgtacc	gccgcggccg	15960
cgcacctctg	accctgccgc	gcgcgcgcta	ccaccgagc	atcgccattt	aaactttcgc	16020

ctgctttgca gatcaatggc cctcacatgc cgccttcgcg ttcccattac gggctaccga	16080
ggaagaaaac cgcgccgtag aaggctggcg gggaacggga tgcgtcgcca ccaccaccgg	16140
cggcggcgcg ccatcagcaa gcggttgggg ggaggcttcc tgcccgcgct gatccccatc	16200
atcgccgcgg cgatcggggc gatccccggc attgcttccg tggcgggtgca ggcctctcag	16260
cgccactgag acacacttgg aaacatcttg taataaacca atggactctg acgctcctgg	16320
tcctgtgatg tgttttcgta gacagatgga agacatcaat ttttcgtccc tggctccgcg	16380
acacggcacg cggccgttca tgggcacctg gagcgacatc ggcaccagcc aactgaacgg	16440
gggcgccttc aattggagca gtctctggag cgggcttaag aatttcgggt ccacgcttaa	16500
aacctatggc agcaaggcgt ggaacagcac cacagggcag gcgtgaggg ataagctgaa	16560
agagcagaac ttccagcaga aggtggtcga tgggctcgcc tcgggcatca acggggtggt	16620
ggacctggcc aaccaggccg tgcagcggca gatcaacagc cgcctggacc cggtgccgcc	16680
cgcgggtccc gtggagatgc cgcaggtgga ggaggagctg cctcccctgg acaagcgggg	16740
cgagaagcga ccccgccccg atgcggagga gacgctgctg acgcacacgg acgagccgcc	16800
cccgtagcag gaggcggtga aactgggtct gccaccacg cggcccatcg cgcccctggc	16860
caccggggtg ctgaaacccg aaaagcccgc gaccctggac ttgcctcctc cccagccttc	16920
ccgcccctct acagtggcta agcccctgcc gccggtggcc gtggcccgcg cgcgaccg	16980
gggcaccgcc cgccctcatg cgaactggca gagcactctg aacagcatcg tgggtctggg	17040
agtgcagagt gtgaagcgcc gccgctgcta ttaaacctac cgtagcgctt aacttgcttg	17100
tctgtgtgtg tatgtattat gtcgccgccg ccgctgtcca ccagaaggag gagtgaagag	17160
gcgcgtcgcc gagttgcaag atggccaccc catcgatgct gcccagtggt gcgtacatgc	17220
acatcgccgg acaggacgct tcggagtacc tgagtccggg tctggtgcag tttgcccgcg	17280
ccacagacac ctacttcagt ctggggaaca agtttaggaa ccccacgggtg gcgcccacgc	17340
acgatgtgac caccgaccgc agccagcggc tgacgctgcg cttcgtgccc gtggaccgcg	17400
aggacaacac ctactcgtac aaagtgcgct acacgctggc cgtgggcgac aaccgcgtgc	17460
tggacatggc cagcacctac tttgacatcc gcggcgtgct ggatcggggc cctagcttca	17520
aaccctactc cggcaccgcc tacaacagtc tggcccccaa gggagcacc aacacttgtc	17580
agtggacata taaagccgat ggtgaaactg ccacagaaaa aacctataca tatggaaatg	17640

cacccgtgca	gggcattaac	atcacaaaag	atggtattca	acttggaact	gacaccgatg	17700
atcagccaat	ctacgcagat	aaaacctatc	agcctgaacc	tcaagtgggt	gatgctgaat	17760
ggcatgacat	cactgggtact	gatgaaaagt	atggaggcag	agctcttaag	cctgatacca	17820
aaatgaagcc	ttgttatgggt	tcttttgcca	agcctactaa	taaagaagga	ggtcaggcaa	17880
atgtgaaaac	aggaacaggc	actactaaag	aatatgacat	agacatggct	ttctttgaca	17940
acagaagtgc	ggctgctgct	ggcctagctc	cagaaattgt	tttgataact	gaaaatgtgg	18000
atttggaac	tccagatacc	catattgtat	acaaagcagg	cacagatgac	agcagctctt	18060
ctattaattt	gggtcagcaa	gccatgccc	acagacctaa	ctacattgggt	ttcagagaca	18120
actttatcgg	gctcatgtac	tacaacagca	ctggcaatat	gggggtgctg	gccggtcagg	18180
cttctcagct	gaatgctgtg	gttgacttgc	aagacagaaa	caccgagctg	tcctaccagc	18240
tcttgcttga	ctctctgggt	gacagaacct	gggtatttcag	tatgtggaat	caggcggtgg	18300
acagctatga	tcctgatgtg	cgcattattg	aaaatcatgg	tgtggaggat	gaacttccca	18360
actattgttt	ccctctggat	gctgttggca	gaacagatac	ttatcaggga	attaaggcta	18420
atggaactga	tcaaaccaca	tggaccaaaag	atgacagtgt	caatgatgct	aatgagatag	18480
gcaagggtaa	tccattcgcc	atggaaatca	acatccaagc	caacctgtgg	aggaacttcc	18540
tctacgcaa	cgtggccctg	tacctgccc	actcttaca	gtacacgccg	gccaatgtta	18600
ccctgcccac	caacaccaac	acctacgatt	acatgaacgg	ccgggtgggtg	gcgccctcgc	18660
tggtggactc	ctacatcaac	atcggggcgc	gctggtcgct	ggatcccatg	gacaacgtga	18720
accccttcaa	ccaccaccgc	aatgcggggc	tgcgctaccg	ctccatgctc	ctgggcaacg	18780
ggcgctacgt	gcccttcac	atccagggtgc	cccagaaatt	tttcgccatc	aagagcctcc	18840
tgctcctgcc	cgggtcctac	acctacgagt	ggaacttccg	caaggacgtc	aacatgatcc	18900
tgagagctc	cctcggaac	gacctgcgca	cggacggggc	ctccatctcc	ttcaccagca	18960
tcaacctcta	cgccaccttc	ttccccatgg	cgcacaacac	ggcctccacg	ctcgaggcca	19020
tgctgcgcaa	cgacaccaac	gaccagtcct	tcaacgacta	cctctcggcg	gccaacatgc	19080
tctaccccat	ccgggccaac	gccaccaacg	tgcccatctc	catcccctcg	cgcaactggg	19140
ccgccttcg	cggctggtcc	ttcacgcgtc	tcaagaccaa	ggagacgccc	tcgctgggct	19200
ccgggttcga	cccctacttc	gtctactcgg	gctccatccc	ctacctcgac	ggcaccttct	19260

acctcaacca	caccttcaag	aaggtctcca	tcaccttcga	ctcctccgtc	agctggcccc	19320
gcaacgaccg	gctcctgacg	cccaacgagt	tcgaaatcaa	gcgcaccgtc	gacggcgagg	19380
gctacaacgt	ggcccagtg	aacatgacca	aggactgggt	cctgggtccag	atgctggccc	19440
actacaacat	cggctaccag	ggcttctacg	tgcccagagg	ctacaaggac	cgcatgtact	19500
ccttcttccg	caacttccag	cccatgagcc	gccagggtgt	ggacgaggtc	aactacaagg	19560
actaccaggc	cgtcaccctg	gcctaccagc	acaacaactc	gggcttcgtc	ggctacctcg	19620
cgcccaccat	gcgccagggc	cagccctacc	ccgccaacta	cccctacccg	ctcatcggca	19680
agagcgccgt	caccagcgct	acccagaaaa	agttcctctg	cgacagggtc	atgtggcgca	19740
tccccttctc	cagcaacttc	atgtccatgg	gcgcgctcac	cgacctcggc	cagaacatgc	19800
tctatgccaa	ctccgcccac	gcgctagaca	tgaatttcga	agtcgacccc	atggatgagt	19860
ccacccttct	ctatgttgct	ttcgaagtct	tcgacgtcgt	ccgagtgcac	cagccccacc	19920
gcggcgcat	cgaggccgtc	tacctgcgca	cccccttctc	ggccggtaac	gccaccacct	19980
aagctcttgc	ttcttgcaag	ccatggccgc	gggctccggc	gagcaggagc	tcagggccat	20040
catccgcgac	ctgggctg	ggccctactt	cctgggcacc	ttcgataagc	gcttcccggg	20100
attcatggcc	ccgcacaagc	tggcctgcgc	catcgtcaac	acggccggcc	gcgagaccgg	20160
gggcgagcac	tggctggcct	tcgcctggaa	cccgcgctcg	aacacctgct	acctcttcga	20220
ccccctcggg	ttctcggacg	agcgcctcaa	gcagatctac	cagttcgagt	acgagggcct	20280
gctgcgccgc	agcgccctgg	ccaccgagga	ccgctgcgtc	accctggaaa	agtccaccca	20340
gaccgtgcag	ggtccgcgct	cggccgcctg	cgggctcttc	tgctgcatgt	tcctgcacgc	20400
cttcgtgcac	tggcccgacc	gccccatgga	caagaacccc	accatgaact	tgctgacggg	20460
ggtgcccac	ggcatgctcc	agtcgcccc	ggtggaaccc	accctgcgcc	gcaaccagga	20520
ggcgctctac	cgcttctca	actcccactc	cgcctacttt	cgctcccacc	gcgcgcgcat	20580
cgagaaggcc	accgccttcg	accgcatgaa	tcaagacatg	taaaccgtgt	gtgtatgtta	20640
aatgtcttta	ataaacagca	ctttcatgtt	acacatgcat	ctgagatgat	ttatttagaa	20700
atcgaaaggg	ttctgccggg	tctcggcatg	gcccgcgggc	agggacacgt	tgcggaactg	20760
gtacttggcc	agccacttga	actcggggat	cagcagtttg	ggcagcgggg	tgctcgggaa	20820
ggagtccgtc	cacagcttcc	gcgtcagttg	cagggcgccc	agcaggtcgg	gcgcggagat	20880

cttgaaatcg cagttgggac ccgcgttctg cgcgcgggag ttgcggtaca cggggttgca	20940
gcactggaac accatcaggg cggggtgctt cacgctcgcc agcaccgtcg cgtcggtgat	21000
gctctccacg tcgaggtcct cggcgttggc catcccgaag ggggtcatct tgcaggtctg	21060
ccttcccatg gtgggcacgc acccgggctt gtggttgcaa tcgcagtgca gggggatcag	21120
catcatctgg gcctggtcgg cgttcatccc cgggtacatg gccttcatga aagcctccaa	21180
ttgcctgaac gcctgctggg ccttggctcc ctcggtgaag aagaccccg caggacttgct	21240
agagaactgg ttggtggcgc acccggcgtc gtgcacgcag cagcgcgcgt cgttgttggc	21300
cagctgcacc acgctgcgc cccagcggtt ctgggtgatc ttggcccgg cggggttctc	21360
cttcagcgcg cgctgcccg tctcgctcgc cacatccatc tcgatcatgt gctccttctg	21420
gatcatggtg gtcccgtgca ggcaccgcag cttgccctcg gcctcggtag acccgtgcag	21480
ccacagcgcg caccggtagc actcccagtt cttgtgggcg atctgggaat gcgcgtgcac	21540
gaagccctgc aggaagcggc ccatcatggt ggtcagggtc ttgttgctag tgaaggtagc	21600
cggaatgccg cggtagctcct cgttgatgta caggtaggcag atgcggcgg acacctcgcc	21660
ctgctcgggc atcagctgga agttggcttt caggtaggtc tccacgcgg agcgggtccat	21720
cagcatagtc atgatttcca tacccttctc ccaggccgag acgatgggca ggctcatagg	21780
gttcttcacc atcatcttag cgctagcagc cgcggccagg gggtagctct cgtccagggt	21840
ctcaaagctc cgcttgccgt ccttctcggg gatccgcacc ggggggtagc tgaagcccac	21900
ggccgccagc tcctcctcgg cctgtctttc gtcctcgctg tcctggctga cgtcctgcag	21960
gaccacatgc ttggtcttgc ggggtttctt cttgggcggc agcggcggcg gagatgttgg	22020
agatggcgag ggggagcgcg agttctcgct caccactact atctcttcct cttcttggtc	22080
cgaggccacg cggcggtagg tatgtctctt cgggggcaga ggcggaggcg acgggctctc	22140
gccgccgcga cttggcggat ggctggcaga gcccttccg cgttcggggg tgcgctccc	22200
gcggcgctct gactgacttc ctccgcggcc ggccattgtg ttctcctagg gaggaacaac	22260
aagcatggag actcagccat cgccaacctc gccatctgcc cccaccgcc acgagaagca	22320
gcagcagcag aatgaaagct taaccgcccc gccgccagc cccgccacct ccgacgcggc	22380
cgtcccagac atgcaagaga tggaggaatc catcgagatt gacctgggct atgtgacgcc	22440
cgcggagcac gaggaggagc tggcagtgcg cttttcacia gaagagatac accaagaaca	22500

gccagagcag gaagcagaga atgagcagag tcaggctggg ctcgagcatg acggcgacta	22560
cctccacctg agcggggggg aggacgcgct catcaagcat ctggcccggc aggccaccat	22620
cgtcaaggat gcgctgctcg accgcaccga ggtgcccctc agcgtggagg agctcagccg	22680
cgcctacgag ttgaacctct tctcgccgcg cgtgcccccc aagcgccagc ccaatggcac	22740
ctgcgagccc aaccgcgcc tcaacttcta cccggtcttc gcggtgcccg aggccctggc	22800
cacctaccac atctttttca agaaccaaaa gatccccgtc tcctgccgcg ccaaccgcac	22860
ccgcgccgac gcccttttca acctgggtcc cggcgcccg cctacctgata tcgcctcctt	22920
ggaagaggtt cccaagatct tcgagggtct gggcagcgac gagactcggg ccgcgaacgc	22980
tctgcaagga gaaggaggag agcatgagca ccacagcgcc ctggtcgagt tggaaggcga	23040
caacgcgcgg ctggcggtgc tcaaacgcac ggtcgagctg acccatttcg cctaccggc	23100
tctgaacctg cccccaaag tcatgagcgc ggtcatggac caggtgctca tcaagcgcg	23160
gtcgcccatc tccgaggacg agggcatgca agactccgag gagggcaagc ccgtggtcag	23220
cgacgagcag ctggcccggg ggctgggtcc taatgctagt cccagagtt tggaagagcg	23280
gcgcaaactc atgatggccg tggctctggt gaccgtggag ctggagtgcc tgcgccgctt	23340
cttcgccgac gcggagacct tgcgcaaggt cgaggagaac ctgcactacc tcttcaggca	23400
cgggttcgtg cgccaggcct gcaagatctc caacgtggag ctgaccaacc tggctctcta	23460
catgggcatc ttgcacgaga accgcctggg gcagaacgtg ctgcacacca ccctgcgcgg	23520
ggaggcccgg cgcgactaca tccgcgactg cgtctacctc tacctctgcc acacctggca	23580
gacgggcatg ggcgtgtggc agcagtgtct ggaggagcag aacctgaaag agctctgcaa	23640
gctcctgcag aagaacctca agggctctgtg gaccgggttc gacgagcgca ccaccgcctc	23700
ggacctggcc gacctcattt tccccgagcg cctcaggctg acgctgcgca acggcctgcc	23760
cgactttatg agccaaagca tgttgcaaaa ctttcgctct ttcatacctcg aacgctccgg	23820
aatcctgccc gccacctgct ccgcgctgcc ctcggacttc gtgccgctga ccttccgcga	23880
gtgccccccg ccgctgtgga gccactgcta cctgctgcgc ctggccaact acctggccta	23940
ccactcggac gtgatcgagg acgtcagcgg cgagggcctg ctcgagtgcc actgccgctg	24000
caacctctgc acgccgcacc gctccctggc ctgcaacccc cagctgctga gcgagacca	24060
gatcatcggc accttcgagt tgcaagggcc cagcgaaggc gagggttcag ccgccaaggg	24120

gggtctgaaa	ctcaccccg	ggctgtggac	ctcggcctac	ttgcgcaagt	tcgtgcccga	24180
ggactacat	cccttcgaga	tcaggttcta	cgaggaccaa	tcccatccgc	ccaaggccga	24240
gctgtcggcc	tgcgatcat	cccagggggc	gacgttgcc	caattgcaag	ccatccagaa	24300
atcccgccaa	gaattcttgc	tgaaaaagg	ccgcggggtc	tacctcgacc	cccagaccgg	24360
tgaggagctc	aaccccggt	tccccagga	tgccccgagg	aaacaagaag	ctgaaagtgg	24420
agctgccgcc	cgtggaggat	ttggaggaag	actgggagaa	cagcagtcag	gcagaggagg	24480
aggagatgga	ggaagactgg	gacagcactc	aggcagagga	ggacagcctg	caagacagtc	24540
tggaggaaga	cgaggaggag	gcagaggagg	aggtggaaga	agcagccgcc	gccagaccgt	24600
cgtcctcggc	gggggagaaa	gcaagcagca	cggataccat	ctccgctccg	ggtcggggtc	24660
ccgctcgacc	acacagtaga	tgggacgaga	ccggacgatt	cccgaacccc	accaccagaa	24720
ccggtaaaga	ggagcggcag	ggatacaagt	cctggcgggg	gcacaaaaac	gccatcgtct	24780
cctgcttgca	ggcctgcggg	ggcaacatct	ccttcacccg	gcgctacctg	ctcttccacc	24840
gcggggtgaa	ctttccccgc	aacatcttgc	attactaccg	tcacctccac	agcccctact	24900
acttccaaga	agaggcagca	gcagcagaaa	aagaccagca	gaaaaccagc	agctagaaaa	24960
tccacagcgg	cggcagcagg	tggactgagg	atcgcggcga	acgagccggc	gcaaaccggg	25020
gagctgagga	accgatctt	tcccaccctc	tatgccatct	tccagcagag	tcgggggcag	25080
gagcaggaac	tgaagtcaa	gaaccgttct	ctgcgctcgc	tcacccgcag	ttgtctgtat	25140
cacaagagcg	aagaccaact	tcagcgcact	ctcaggagcg	ccgaggctct	cttcaacaag	25200
tactgcgcgc	tcactcttaa	agagtagccc	gcgcccgcgc	agtcgcagaa	aaaggcggga	25260
attacgtcac	ctgtgccctt	cgccttagcc	gcctccaccc	atcatcatga	gcaaagagat	25320
tcccacgcct	tacatgtgga	gctaccagcc	ccagatgggc	ctggccgccg	gtgccgccc	25380
ggactactcc	acccgcatga	attggctcag	cgccggggcc	gcgatgatct	cacgggtgaa	25440
tgacatccgc	gcccaccgaa	accagatact	cctagaacag	tcagcgtca	ccgccacgcc	25500
ccgcaatcac	ctcaatccgc	gtaattggcc	cgccgccctg	gtgtaccagg	aaattcccca	25560
gcccacgacc	gtactacttc	cgcgagacgc	ccaggccgaa	gtccagctga	ctaactcagg	25620
tgtccagctg	gcgggcggcg	ccaccctgtg	tcgtcaccgc	cccgctcagg	gtataaagcg	25680
gctggtgatc	cggggcagag	gcacacagct	caacgacgag	gtggtgagct	cttcgctggg	25740

tctgcgacct gacggagtct tccaactcgc cggatcgggg agatcttcct tcacgcctcg	25800
tcaggccgtc ctgactttgg agagttcgtc ctcgcagccc cgctcgggtg gcatcggcac	25860
tctccagttc gtggaggagt tcactccctc ggtctacttc aacccttct ccggctcccc	25920
cggccactac ccggacgagt tcatcccgaa cttcgacgcc atcagcgagt cgggtggacgg	25980
ctacgattga aactaatcac ccccttatcc agtgaaataa agatcatatt gatgatgatt	26040
ttacagaaat aaaaaataat catttgattt gaaataaaga tacaatcata ttgatgattt	26100
gagttaaca aaaaaataaa gaatcactta cttgaaatct gataccaggt ctctgtccat	26160
gttttctgcc aacaccactt cactccctc ttcccagctc tgggtactgca ggccccggcg	26220
ggctgcaaac ttctccaca cgctgaaggg gatgtcaaat tcctcctgtc cctcaatctt	26280
cattttatct tctatcagat gtccaaaaag cgcgtccggg tggatgatga cttcgacccc	26340
gtctaccct acgatgcaga caacgcaccg accgtgccct tcatcaaccc ccccttcgtc	26400
tcttcagatg gattccaaga gaagcccctg ggggtgttgt ccctgcgact ggccgacccc	26460
gtcaccacca agaacgggga aatcacctc aagctgggag agggggtgga cctcgattcc	26520
tcgggaaaac tcatctcaa cacggccacc aaggccgccg cccctctcag tttttccaac	26580
aacaccattt cccttaacat ggatcacccc ttttacacta aagatggaaa attatcctta	26640
caagtttctc caccattaaa tatactgaga acaagcattc taaacacact agctttaggt	26700
tttggatcag gtttaggact ccgtggctct gccttggcag tacagttagt ctctccactt	26760
acatttgata ctgatggaaa cataaagctt accttagaca gaggtttgca tgttacaaca	26820
ggagatgcaa ttgaaagcaa cataagctgg gctaaagggt taaaatttga agatggagcc	26880
atagcaacca acattggaaa tgggttagag tttggaagca gtagtacaga aacaggtgtt	26940
gatgatgctt acccaatcca agttaaactt ggatctggcc ttagctttga cagtacagga	27000
gccataatgg ctggtaacaa agaagacgat aaactcactt tgttgacaac acctgatcca	27060
tcaccaaact gtcaaatact cgcagaaaat gatgcaaac taacactttg cttgactaaa	27120
tgtggtagtc aaatactggc cactgtgtca gtcttagttg taggaagtgg aaacctaaac	27180
cccattactg gcaccgtaag cagtgtcag gtgtttctac gttttgatgc aaacggtgtt	27240
cttttaacag aacattctac actaaaaaaaa tactgggggt ataggcaggg agatagcata	27300
gatggcactc catataccaa tgctgtagga ttcatgcccc atttaaaagc ttatccaaag	27360

tcacaaagtt ctactactaa aaataatata gtagggcaag tatacatgaa tggagatgtt	27420
tcaaaaccta tgcttctcac tataaccctc aatgggtactg atgacagcaa cagtacatat	27480
tcaatgtcat ttccatacac ctggactaat ggaagctatg ttggagcaac atttggggct	27540
aactcttata ctttctcata catcgcccaa gaatgaacac tgtatcccac cctgcatgcc	27600
aacccttccc accccactct gtggaacaaa ctctgaaaca caaaataaaa taaagttcaa	27660
gtgttttatt gattcaacag ttttacagga ttcgagcagt tatttttcct ccaccctccc	27720
aggacatgga atacaccacc ctctccccc gcacagcctt gaacatctga atgccattgg	27780
tgatggacat gcttttggtc tccacgttcc acacagtttc agagcgagcc agtctcgggt	27840
cggtcaggga gatgaaaccc tccgggcact cccgcatctg cacctcacag ctcaacagct	27900
gaggattgtc ctcggtggtc gggatcacgg ttatctggaa gaagcagaag agcggcggtg	27960
ggaatcatag tccgcgaacg ggatcggccg gtggtgtcgc atcaggcccc gcagcagtcg	28020
ctgccgccgc cgctccgtca agctgctgct caggggggtcc ggggtccaggg actccctcag	28080
catgatgccc acggccctca gcatcagtcg tctggtgcgg cgggcgcagc agcgcagtcg	28140
gatctcgctc aggtcgctgc agtacgtgca acacagaacc accaggttgt tcaacagtcc	28200
atagttcaac acgctccagc cgaaactcat cgcgggaagg atgctacca cgtggccgtc	28260
gtaccagatc ctcaggtaaa tcaagtgggtg cccctccag aacacgctgc ccacgtacat	28320
gatctccttg ggcattgtggc ggttcaccac ctcccgttac cacatcacc cttggttgaa	28380
catgcagccc cggatgatcc tgcggaacca cagggccagc accgccccgc ccgcatgca	28440
gcgaagagac cccgggtccc ggcaatggca atggaggacc caccgctcgt acccggtgat	28500
catctgggag ctgaacaagt ctatgttggc acagcacagg catatgctca tgcattctt	28560
cagcactctc aactcctcgg ggggtcaaac catatcccag ggcacgggga actcttgcat	28620
gacagcgaac cccgcagaac agggcaatcc tcgcacagaa cttacattgt gcatggacag	28680
ggtatcgcaa tcaggcagca ccgggtgatc ctccaccaga gaagcgcggg tctcggctc	28740
ctcacagcgt ggtaaggggg ccggccgata cgggtgatgg cgggacgcgg ctgatcgtgt	28800
tcgcgaccgt gtcattgatgc agttgctttc ggacattttc gtacttgctg tagcagaacc	28860
tgggtccgggc gctgcacacc gatcgccggc ggcggtctcg gcgcttgga cgctcgggtgt	28920
tgaaattgta aaacagccac tctctcagac cgtgcagcag atctagggcc tcaggagtga	28980

tgaagatccc atcatgcctg atggctctga tcacatcgac caccgtggaa tgggccagac	29040
ccagccagat gatgcaattt tgttgggttt cggtgacggc gggggaggga agaacaggaa	29100
gaaccatgat taacttttaa tccaaacggt ctcgaggtac ttcaaaatga agatcgcgga	29160
gatggcacct ctcgcccccg ctgtgttggt ggaaaataac agccagggtca aagggtgatac	29220
ggttctcgag atgttccacg gtggcttcca gcaaagcctc cacgcgaca tccagaaaca	29280
agacaatagc gaaagcggga gggttctcta attcctcaat catcatgtta cactcctgca	29340
ccatccccag ataattttca tttttccagc cttgaatgat tcgaactagt tcctgaggta	29400
aatccaagcc agccatgata aagagctcgc gcagagcgcc ctccaccggc attcttaagc	29460
acaccctcat aattccaaga tattctgctc ctggttcacc tgcagcagat tgacaagcgg	29520
aatatcaaaa tctctgccgc gatccctgag ctctccctc agcaataact gtaagtactc	29580
tttcatatcc tctccgaaat ttttagccat aggaccacca ggaataagat tagggcaagc	29640
cacagtacag ataaaccgaa gtcctcccca gtgagcattg ccaaatgcaa gactgctata	29700
agcatgctgg ctagaccggg tgatatcttc cagataactg gacagaaaaat cgcccaggca	29760
atttttaaga aaatcaacaa aagaaaaatc ctccagggtg acgttttagag cctcgggaac	29820
aacgatgaag taaatgcaag cgggtgcgttc cagcatgggt agttagctga tctgtagaaa	29880
aaacaaaaat gaacattaaa ccatgctagc ctggcgaaca ggtgggtaaa tcgttctctc	29940
cagcaccagg caggccacgg ggtctccggc gcgaccctcg taaaaattgt cgctatgatt	30000
gaaaaccatc acagagagac gttcccgggt gccggcgtga atgattcgac aagatgaata	30060
cacccccgga acattggcgt ccgcgagtga aaaaaagcgc ccgaggaagc aataaggcac	30120
tacaatgctc agtctcaagt ccagcaaagc gatgccatgc ggatgaagca caaaattctc	30180
aggtgcgtac aaaatgtaat tactcccctc ctgcacaggc agcaaagccc ccgatccctc	30240
caggtacaca taaaagcct cagcgtccat agcttaccga gcagcagcac acaacaggcg	30300
caagagtcag agaaaggctg agctctaacc tgtccacccg ctctctgctc aatatatagc	30360
ccagatctac actgacgtaa aggccaaagt ctaaaaatac ccgccaata atcacacacg	30420
cccagcacac gccagaaac cggtgacaca ctcaaaaaaa tacgcgact tcctcaaacg	30480
cccaaaactg ccgtcatttc cgggttcca cgctacgtca taaaacacg actttcaa	30540
tccgtcgacc gttaaaaacg tcaccgccc cgcccctaac ggctgcccgt ctctcagcca	30600

atcagcgcgc cgcacccca aattcaaaca cctcatttgc atattaacgc gcacaaaaag	30660
tttgaggtat attattgatg atgg	30684

<210> 14
 <211> 8602
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 14	
atgggcggcg catgagagaa gccagacca attacctacc caaaatggag aaagttcacg	60
ttgacatcga ggaagacagc ccattcctca gagctttgca gcggagcttc ccgcagtttg	120
aggtagaagc caagcaggtc actgataatg accatgctaa tgccagagcg ttttcgcatc	180
tggcttcaaa actgatcgaa acggaggtgg acccatccga cacgatcctt gacattggaa	240
gtgcgcccgc ccgcagaatg tattctaagc acaagtatca ttgtatctgt ccgatgagat	300
gtgcggaaga tccggacaga ttgtataagt atgcaactaa gctgaagaaa aactgtaagg	360
aaataactga taaggaattg gacaagaaaa tgaaggagct cgccgccgtc atgagcgacc	420
ctgacctgga aactgagact atgtgcctcc acgacgacga gtcgtgtcgc tacgaagggc	480
aagtcgctgt ttaccaggat gtatacgcgg ttgacggacc gacaagtctc tatcaccaag	540
ccaataaggg agttagagtc gcctactgga taggctttga caccaccctt tttatgttta	600
agaacttggc tggagcatat ccatcatact ctaccaactg ggccgacgaa accgtgttaa	660
cggctcgtaa cataggccta tgcagctctg acgttatgga gcggtcacgt agagggatgt	720
ccattcttag aaagaagtat ttgaaaccat ccaacaatgt tctattctct gttggctcga	780
ccatctacca cgagaagagg gacttactga ggagctggca cctgccgtct gtatttcact	840
tacgtggcaa gcaaaattac acatgtcggg gtgagactat agttagttgc gacgggtacg	900
tcgttaaaaag aatagctatc agtccaggcc tgtatgggaa gccttcaggc tatgctgcta	960
cgatgcaccg cgagggatc ttgtgctgca aagtgacaga cacattgaac ggggagaggg	1020
tctcttttcc cgtgtgcacg tatgtgccag ctacattgtg tgaccaaatg actggcatac	1080
tggcaacaga tgtcagtgcg gacgacgcgc aaaaactgct gggtgggctc aaccagcgta	1140
tagtcgtcaa cggtcgcacc cagagaaaca ccaataccat gaaaaattac cttttgccc	1200

tagtggccca	ggcatttgct	aggtgggcaa	aggaatataa	ggaagatcaa	gaagatgaaa	1260
ggccactagg	actacgagat	agacagttag	tcatgggggtg	ttgttgggct	tttagaaggc	1320
acaagataac	atctatttat	aagcgcccgg	atacccaaac	catcatcaaa	gtgaacagcg	1380
atttcactc	attcgtgctg	cccaggatag	gcagtaacac	attggagatc	gggctgagaa	1440
caagaatcag	gaaaatgtta	gaggagcaca	aggagccgtc	acctctcatt	accgccgagg	1500
acgtacaaga	agctaagtgc	gcagccgatg	aggctaagga	ggtgcgtgaa	gccgaggagt	1560
tgcgcgcagc	tctaccacct	ttggcagctg	atgttgagga	gcccactctg	gaagccgatg	1620
tcgacttgat	gttacaagag	gctggggccg	gctcagtgga	gacacctcgt	ggcttgataa	1680
aggttaccag	ctacgctggc	gaggacaaga	tcggctctta	cgctgtgctt	tctccgcagg	1740
ctgtactcaa	gagtgaaaaa	ttatcttgca	tccaccctct	cgctgaacaa	gtcatagtga	1800
taacacactc	tggccgaaaa	gggcgttatg	ccgtggaacc	ataccatggg	aaagtagtgg	1860
tgccagaggg	acatgcaata	cccgtccagg	actttcaagc	tctgagtgaa	agtgccacca	1920
tttgttacia	cgaacgtgag	ttcgtaaaca	ggtacctgca	ccatattgcc	acacatggag	1980
gagcgtgaa	cactgatgaa	gaatattaca	aaactgtcaa	gcccagcgag	cacgacggcg	2040
aatacctgta	cgacatcgac	aggaaacagt	gcgtcaagaa	agaactagtc	actgggctag	2100
ggctcacagg	cgagctgggtg	gacctccct	tccatgaatt	cgcttacgag	agtctgagaa	2160
cacgaccagc	cgctccttac	caagtaccaa	ccataggggt	gtatggcgtg	ccaggatcag	2220
gcaagtctgg	catcattaaa	agcgcagtca	caaaaaaga	tctagtgggtg	agcgccaaga	2280
aagaaaactg	tgcagaaatt	ataagggacg	tcaagaaaat	gaaagggtg	gacgtcaatg	2340
ccagaactgt	ggactcagtg	ctcttgaatg	gatgcaaaca	ccccgtagag	accctgtata	2400
ttgacgaagc	ttttgcttgt	catgcaggta	ctctcagagc	gctcatagcc	attataagac	2460
ctaaaaaggc	agtgctctgc	ggggatccca	aacagtgcgg	tttttttaac	atgatgtgcc	2520
tgaaagtgca	ttttaaccac	gagatttgca	cacaagtctt	ccacaaaagc	atctctcgcc	2580
gttgactaa	atctgtgact	tcggtcgtct	caaccttggt	ttacgacaaa	aaaatgagaa	2640
cgacgaatcc	gaaagagact	aagattgtga	ttgacactac	cggcagtacc	aaacctaagc	2700
aggacgatct	cattctcact	tgtttcagag	ggtgggtgaa	gcagttgcaa	atagattaca	2760
aaggcaacga	aataatgacg	gcagctgcct	ctcaagggt	gacccgtaaa	ggtgtgtatg	2820

ccgttcggta caaggtgaat gaaaatcctc tgtacgcacc cacctcagaa catgtgaacg	2880
tcctactgac ccgcacggag gaccgcatcg tgtggaaaac actagccggc gacccatgga	2940
taaaaacact gactgccaag taccctggga atttcactgc cacgatagag gagtggcaag	3000
cagagcatga tgccatcatg aggcacatct tggagagacc ggaccctacc gacgtcttcc	3060
agaataaggc aaacgtgtgt tgggccaagg ctttagtgcc ggtgctgaag accgctggca	3120
tagacatgac cactgaacaa tggaacactg tggattattt tgaaacggac aaagctcact	3180
cagcagagat agtattgaac caactatgcg tgaggttctt tggactcgat ctggactccg	3240
gtctattttc tgcaccact gttccgttat ccattaggaa taatcactgg gataactccc	3300
cgtcgcctaa catgtacggg ctgaataaag aagtgggtccg tcagctctct cgcaggtacc	3360
cacaactgcc tcgggcagtt gccactggaa gagtctatga catgaacact ggtacactgc	3420
gcaattatga tccgcgcata aacctagtac ctgtaaacag aagactgcct catgctttag	3480
tcctccacca taatgaacac ccacagagtg acttttcttc attcgtcagc aaattgaagg	3540
gcagaactgt cctgggtggtc ggggaaaagt tgtccgtccc aggcaaatg gttgactggt	3600
tgtcagaccg gcctgaggct accttcagag ctcggctgga tttaggcatc ccaggtgatg	3660
tgcccaaata tgacataata tttgttaatg tgaggacccc atataaatac catcactatc	3720
agcagtgtga agaccatgcc attaagctta gcatgttgac caagaaagct tgtctgcatc	3780
tgaatcccgg cggaacctgt gtcagcatag gttatggtta cgctgacagg gccagcgaaa	3840
gcatcattgg tgctatagcg cggcagttca agttttcccg ggtagtcaaa ccgaaatcct	3900
cacttgaaga gacggaagtt ctgtttgtat tcattgggta cgatcgcaag gcccgtacgc	3960
acaatcctta caagctttca tcaaccttga ccaacattta tacaggttcc agactccacg	4020
aagccggatg tgcaccctca tatcatgtgg tgcgagggga tattgccacg gccaccgaag	4080
gagtgattat aaatgctgct aacagcaaag gacaacctgg cggaggggtg tgcggagcgc	4140
tgtataagaa attcccggaa agcttcgatt tacagccgat cgaagtagga aaagcgcgac	4200
tgggtcaaagg tgcagctaaa catatcattc atgccgtagg accaaacttc aacaaagttt	4260
cggagggtga aggtgacaaa cagttggcag aggcttatga gtccatcgct aagattgtca	4320
acgataacaa ttacaagtca gtagcgattc cactgttgtc caccggcatc ttttccggga	4380
acaaagatcg actaaccaa tcattgaacc atttgctgac agcttttagac accactgatg	4440

cagatgtagc catatactgc agggacaaga aatgggaaat gactctcaag gaagcagtgg	4500
ctaggagaga agcagtggag gagatatgca tatccgacga ctcttcagtg acagaacctg	4560
atgcagagct ggtgagggtg catccgaaga gttctttggc tggaaggaag ggctacagca	4620
caagcgatgg caaaactttc tcatatatttg aagggaacca gtttcaccag gcggccaagg	4680
atatagcaga aattaatgcc atgtggcccg ttgcaacgga ggccaatgag caggtatgca	4740
tgtatatcct cggagaaagc atgagcagta ttaggtcgaa atgccccgtc gaagagtcgg	4800
aagcctccac accacctagc acgctgcctt gcttgtgcat ccatgccatg actccagaaa	4860
gagtacagcg cctaaaagcc tcacgtccag aacaaattac tgtgtgctca tcctttccat	4920
tgccgaagta tagaatcact ggtgtgcaga agatccaatg ctcccagcct atattgttct	4980
caccgaaagt gcctgcgtat attcatcaa ggaagtatct cgtggaaaca ccaccggtag	5040
acgagactcc ggagccatcg gcagagaacc aatccacaga ggggacacct gaacaaccac	5100
cacttataac cgaggatgag accaggacta gaacgcctga gccgatcatc atcgaagagg	5160
aagaagagga tagcataagt ttgctgtcag atggcccgac ccaccagggtg ctgcaagtcg	5220
aggcagacat tcacgggccg ccctctgtat ctagctcatc ctggtccatt cctcatgcat	5280
ccgactttga tgtggacagt ttatccatac ttgacaccct ggaggggagct agcgtgacca	5340
gcgggggcaac gtcagccgag actaactctt acttcgcaaa gagtatggag tttctggcgc	5400
gaccggtgcc tgcgccctga acagtattca ggaaccctcc acatcccgtc ccgcgcaaca	5460
gaacaccgtc acttgcaccc agcagggcct gctcgagaac cagcctagtt tccaccccg	5520
caggcgtgaa tagggtgatc actagagagg agctcgaggc gcttaccgct tcacgcactc	5580
ctagcaggtc ggtctcgaga accagcctgg tctccaaccc gccaggcgta aatagggtga	5640
ttacaagaga ggagtttgag gcgttcgtag cacaacaaca atgacgggtt gatgcgggtg	5700
catacatctt ttcctccgac accggtcaag ggcatttaca acaaaaatca gtaaggcaaa	5760
cgggtgctatc cgaagtgggtg ttggagagga ccgaattgga gatttcgtat gccccgcgcc	5820
tcgaccaaga aaaagaagaa ttactacgca agaaattaca gttaaatccc acacctgcta	5880
acagaagcag ataccagtcc aggaagggtg agaacatgaa agccataaca gctagacgta	5940
ttctgcaagg cctagggcat tatttgaagg cagaaggaaa agtggagtgc taccgaacct	6000
tgcacctgt tcctttgtat tcacttagtg tgaaccgtgc cttttcaagc cccaaggctg	6060

cagtggaagc ctgtaacgcc atgttgaaag agaactttcc gactgtggct tcttactgta	6120
ttattccaga gtacgatgcc tatttggaca tggttgacgg agcttcatgc tgcttagaca	6180
ctgccagttt ttgccctgca aagctgcgca gctttccaaa gaaacactcc tatttggaac	6240
ccacaatacg atcggcagtg ctttcagcga tccagaacac gctccagaac gtcctggcag	6300
ctgccacaaa aagaaattgc aatgtcacgc aaatgagaga attgcccgtg ttggattcgg	6360
cggcctttta tgtggaatgc ttcaagaaat atgcgtgtaa taatgaatat tgggaaacgt	6420
ttaaagaaaa ccccatcagg cttactgaag aaaacgtggg aaattacatt accaaattaa	6480
aaggaccaa agctgctgct ctttttgcga agacacataa tttgaatatg ttgcaggaca	6540
taccaatgga caggtttgta atggacttaa agagagacgt gaaagtgact ccaggaacaa	6600
aacatactga agaacggccc aaggtacagg tgatccaggc tgccgatccg ctagcaacag	6660
cgtatctgtg cggaatccac cgagagctgg ttaggagatt aaatgcggtc ctgcttccga	6720
acattcatac actgtttgat atgtcggctg aagactttga cgctattata gccgagcact	6780
tccagcctgg ggattgtgtt ctggaaactg acatcgcgtc gtttgataaa agtgaggacg	6840
acgccatggc tctgaccgcg ttaatgattc tggaagactt aggtgtggac gcagagctgt	6900
tgacgctgat tgaggcggct ttcggcgaaa tttcatcaat acatttgccc actaaaacta	6960
aatttaaat cggagccatg atgaaatctg gaatgttcct cacactgttt gtgaacacag	7020
tcattaacat tgtaatcgca agcagagtgt tgagagaacg gctaaccgga tcaccatgtg	7080
cagcattcat tggagatgac aatatcgtga aaggagtcaa atcggacaaa ttaatggcag	7140
acagggtgcg cacctggttg aatatggaag tcaagattat agatgctgtg gtgggcgaga	7200
aagcgcctta tttctgtgga gggtttattt tgtgtgactc cgtgaccggc acagcgtgcc	7260
gtgtggcaga cccctaaaa aggctgttta agcttggtcaa acctctggca gcagacgatg	7320
aacatgatga tgacaggaga agggcattgc atgaagagtc aacacgctgg aaccgagtgg	7380
gtattctttc agagctgtgc aaggcagtag aatcaaggta tgaaaccgta ggaacttcca	7440
tcatagttat ggccatgact actctagcta gcagtgttaa atcattcagc tacctgagag	7500
gggcccctat aactctctac ggctaacctg aatggactac gactctagaa tagtctttta	7560
ttaaagtccg ccatatgagg ccaccatgca gatcttcgtg aagaccctga ccggcaagac	7620
catcacccta gaggtggagc ccagtgcac catcgagaac gtgaaggcca agatccagga	7680

taaagagggc atccccctg accagcagag gctgatcttt gccggcaagc agctggaaga	7740
tggccgcacc ctctctgatt acaacatcca gaaggagtca accctgcacc tggtccttcg	7800
cctgagaggt ggcgctgctt acagtataat caactttgaa aaactggctg cttacggcat	7860
cctgggcttt gtgtttacac tggctgccta cctgctgttt ggctatcctg tgtacgtggc	7920
cgcttatgga ctgtgtaccc tgggtggccat gctggctgct tacaatctgg tgcctatggt	7980
ggccacagtg gccgcctatt gtcttggcgg actgctgaca atggtggcag cctacagccc	8040
gagctatgcg tatcatcagt ttgcagccta cggcccagga ccaggcgcta aatttgtggc	8100
tgcctggaca ctgaaagccg ccgctggacc aggtcctgga cagtacatca aggccaacag	8160
caagttcatc ggcacacccg aactcggccc aggaccaggc tatccctacg atgtgcctga	8220
ttacgcctga tagtgatgat tcgaacggcc gtatcacgcc caaacattta cagccgcggt	8280
gtcaaaaacc gcgtggacgt ggttaacatc cctgctggga ggatcagccg taattattat	8340
aattggcttg gtgctggcta ctattgtggc catgtacgtg ctgaccaacc agaaacataa	8400
ttgaatacag cagcaattgg caagctgctt acatagaact cgcggcgatt ggcatgccgc	8460
cttaaaattt ttattttatt ttttcttttc ttttccgaat cggattttgt ttttaattat	8520
tcaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa	8580
aaaaaaaaaa aaaaaaaaaa aa	8602

<210> 15

<211> 9595

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 15

atgggcggcg catgagagaa gcccagacca attacctacc caaaatggag aaagttcacg	60
ttgacatcga ggaagacagc ccattcctca gagctttgca gcggagcttc ccgcagtttg	120
aggtagaagc caagcaggtc actgataatg accatgctaa tgccagagcg ttttcgcatc	180
tggcttcaaa actgatcgaa acggagggtg acccatccga cacgatcctt gacattggaa	240
gtgcgcccgc ccgcagaatg tattctaagc acaagtatca ttgtatctgt ccgatgagat	300

gtgcggaaga tccggacaga ttgtataagt atgcaactaa gctgaagaaa aactgtaagg	360
aaataactga taaggaattg gacaagaaaa tgaaggagct cgccgccgtc atgagcgacc	420
ctgacctgga aactgagact atgtgcctcc acgacgacga gtcgtgtcgc tacgaagggc	480
aagtcgctgt ttaccaggat gtatacgcggt ttgacggacc gacaagtctc tatcaccaag	540
ccaataaggg agtttagagtc gcctactgga taggctttga caccaccctt tttatgttta	600
agaacttggc tggagcatat ccatcatact ctaccaactg ggccgacgaa accgtgttaa	660
cggctcgtaa cataggccta tgcagctctg acgttatgga gcgggtcacgt agaggggatgt	720
ccattcttag aaagaagtat ttgaaacat ccaacaatgt tctattctct gttggctcga	780
ccatctacca cgagaagagg gacttactga ggagctggca cctgccgtct gtatttcact	840
tacgtggcaa gcaaaattac acatgtcggg gtgagactat agttagttgc gacgggtacg	900
tcgttaaaag aatagctatc agtccaggcc tgtatgggaa gccttcaggc tatgctgcta	960
cgatgcaccg cgagggattc ttgtgctgca aagtgcaga cacattgaac ggggagaggg	1020
tctcttttcc cgtgtgcacg tatgtgccag ctacattgtg tgaccaaattg actggcatac	1080
tggcaacaga tgtcagtgcg gacgacgcgc aaaaactgct gggtgggctc aaccagcgta	1140
tagtcgtcaa cggtcgcacc cagagaaaca ccaataccat gaaaaattac cttttgcccg	1200
tagtggccca ggcatttgct aggtgggcaa aggaatataa ggaagatcaa gaagatgaaa	1260
ggccactagg actacgagat agacagttag tcatgggggtg ttgttgggct ttagaaggc	1320
acaagataac atctatttat aagcgcccgg atacccaaac catcatcaaa gtgaacagcg	1380
atttcactc attcgtgctg cccaggatag gcagtaacac attggagatc gggctgagaa	1440
caagaatcag gaaaatgtta gaggagcaca aggagccgtc acctctcatt accgccgagg	1500
acgtacaaga agctaagtgc gcagccgatg aggctaagga ggtgcgtgaa gccgaggagt	1560
tgcgcgcagc tctaccacct ttggcagctg atgttgagga gcccactctg gaagccgatg	1620
tcgacttgat gttacaagag gctggggccg gctcagtgga gacacctcgt ggcttgataa	1680
aggttaccag ctacgctggc gaggacaaga tcggctctta cgctgtgctt tctccgcagg	1740
ctgtactcaa gagtgaaaaa ttatcttgca tccaccctct cgctgaacaa gtcatagtga	1800
taacacactc tggccgaaaa gggcgttatg ccgtggaacc ataccatggg aaagtagtgg	1860
tgccagaggg acatgcaata cccgtccagg actttcaagc tctgagtga agtgccacca	1920

ttgtgtacaa cgaacgtgag ttcgtaaaca ggtacctgca ccatattgcc acacatggag	1980
gagcgcgtgaa cactgatgaa gaatattaca aaactgtcaa gcccagcgag cacgacggcg	2040
aatacctgta cgacatcgac aggaaacagt gcgtcaagaa agaactagtc actgggctag	2100
ggctcacagg cgagctgggtg gatcctccct tccatgaatt cgcctacgag agtctgagaa	2160
cacgaccagc cgctccttac caagtaccaa ccataggggt gtatggcgtg ccaggatcag	2220
gcaagtctgg catcattaaa agcgcagtca ccaaaaaaga tctagtgggtg agcgccaaga	2280
aagaaaactg tgcagaaatt ataagggacg tcaagaaaat gaaagggtg gacgtcaatg	2340
ccagaactgt ggactcagtg ctcttgaatg gatgcaaaca ccccgtagag accctgtata	2400
ttgacgaagc ttttgcttgt catgcaggta ctctcagagc gctcatagcc attataagac	2460
ctaaaaaggc agtgctctgc ggggatccca aacagtgcgg tttttttaac atgatgtgcc	2520
tgaaagtgca ttttaaccac gagatttgca cacaagtctt ccacaaaagc atctctcgcc	2580
gttgactaa atctgtgact tcggtcgtct caaccttggt ttacgacaaa aaaatgagaa	2640
cgacgaatcc gaaagagact aagattgtga ttgacactac cggcagtacc aaacctaagc	2700
aggacgatct cattctcact tgtttcagag ggtgggtgaa gcagttgcaa atagattaca	2760
aaggcaacga aataatgacg gcagctgcct ctcaagggtg gacccgtaaa ggtgtgtatg	2820
ccgttcggta caaggtgaat gaaaatcctc tgtacgcacc cacctcagaa catgtgaacg	2880
tcctactgac ccgcacggag gaccgcatcg tgtggaaaac actagccggc gacccatgga	2940
taaaaaact gactgccaag taccctggga atttcactgc cacgatagag gagtggcaag	3000
cagagcatga tgccatcatg aggcacatct tggagagacc ggaccctacc gacgtcttcc	3060
agaataaggc aaacgtgtgt tgggccaagg ctttagtgcc ggtgctgaag accgctggca	3120
tagacatgac cactgaacaa tggaacactg tggattatct tgaaacggac aaagctcact	3180
cagcagagat agtattgaac caactatgcg tgaggttctt tggactcgat ctggactccg	3240
gtctatcttc tgcaccact gttccgttat ccattaggaa taatcactgg gataactccc	3300
cgtcgcctaa catgtacggg ctgaataaag aagtgggtccg tcagctctct cgcaggtacc	3360
cacaactgcc tcgggcagtt gccactggaa gagtctatga catgaacact ggtacactgc	3420
gcaattatga tccgcgcata aacctagtac ctgtaaacag aagactgcct catgctttag	3480
tcctccacca taatgaacac ccacagagtg acttttcttc attcgtcagc aaattgaagg	3540

gcagaactgt cctggtggtc ggggaaaagt tgtccgtccc aggcaaaatg gttgactggt	3600
tgtcagaccg gcctgaggct accttcagag ctcggtgga tttaggcatc ccaggatgatg	3660
tgcccaaata tgacataata tttgttaatg tgaggacccc atataaatac catcactatc	3720
agcagtgtga agaccatgcc attaatgctta gcatgttgac caagaaagct tgtctgcatc	3780
tgaatcccgg cggaacctgt gtcagcatag gttatgggta cgctgacagg gccagcgaaa	3840
gcatcattgg tgctatagcg cggcagttca agttttcccg ggtatgcaaa ccgaaatcct	3900
cacttgaaga gacggaagtt ctgtttgtat tcattgggta cgatcgcaag gcccgtacgc	3960
acaatcctta caagctttca tcaaccttga ccaacattta tacagggttc agactccacg	4020
aagccggatg tgcaccctca tatcatgtgg tgcgagggga tattgccacg gccaccgaag	4080
gagtgattat aaatgctgct aacagcaaag gacaacctgg cggaggggtg tgcggagcgc	4140
tgtataagaa attcccggaa agcttcgatt tacagccgat cgaagtagga aaagcgcgac	4200
tggtcaaagg tgcagctaaa catatcattc atgccgtagg accaaacttc aacaaagttt	4260
cggaggttga aggtgacaaa cagttggcag aggcttatga gtccatcgct aagattgtca	4320
acgataacaa ttacaagtca gtagcgattc cactgttgct caccggcatc ttttccggga	4380
acaaagatcg actaacccaa tcattgaacc atttgctgac agcttttagac accactgatg	4440
cagatgtagc catatactgc agggacaaga aatgggaaat gactctcaag gaagcagtgg	4500
ctaggagaga agcagtggag gagatatgca tatccgacga ctcttcagt acagaacctg	4560
atgcagagct ggtgaggggtg catccgaaga gttctttggc tggaagggaag ggctacagca	4620
caagcgatgg caaaactttc tcatatattg aagggaacca gtttcaccag gcggccaagg	4680
atatagcaga aattaatgcc atgtggcccg ttgcaacgga ggccaatgag caggtatgca	4740
tgtatatcct cggagaaagc atgagcagta ttaggtcgaa atgccccgtc gaagagtcgg	4800
aagcctccac accacctagc acgctgcctt gcttgtgcat ccatgccatg actccagaaa	4860
gagtacagcg cctaaaagcc tcacgtccag aacaaattac tgtgtgctca tcctttccat	4920
tgccgaagta tagaatcact ggtgtgcaga agatccaatg ctcccagcct atattgttct	4980
caccgaaagt gcctgcgtat attcatcaa ggaagtatct cgtggaaaca ccaccggtag	5040
acgagactcc ggagccatcg gcagagaacc aatccacaga ggggacacct gaacaaccac	5100
cacttataac cgaggatgag accaggacta gaacgcctga gccgatcatc atcgaagagg	5160

aagaagagga tagcataagt ttgctgtcag atggcccgac ccaccagggtg ctgcaagtcg	5220
aggcagacat tcacggggccg ccctctgtat ctagctcatc ctggtccatt cctcatgcat	5280
ccgactttga tgtggacagt ttatccatac ttgacaccct ggagggagct agcgtgacca	5340
gcggggcaac gtcagccgag actaactctt acttcgcaaa gagtatggag tttctggcgc	5400
gaccggtgcc tgcgcctcga acagtattca ggaaccctcc acatcccgt cgcgcacaa	5460
gaacaccgtc acttgcaccc agcagggcct gctcgagaac cagcctagtt tccaccccgc	5520
caggcgtgaa tagggatgac actagagagg agctcgaggc gcttaccctg tcacgcactc	5580
ctagcaggtc ggtctcgaga accagcctgg tctccaaccc gccaggcgta aatagggtga	5640
ttacaagaga ggagtttgag gcgttcgtag cacaacaaca atgacggttt gatgcgggtg	5700
catacatctt ttcctccgac accgggtcaag ggcatttaca acaaaaatca gtaaggcaaa	5760
cgggtgctatc cgaagtgggtg ttggagagga ccgaattgga gatttcgtat gccccgcgcc	5820
tcgaccaaga aaaagaagaa ttactacgca agaaattaca gttaaattccc acacctgcta	5880
acagaagcag ataccagtcc aggaagggtg agaacatgaa agccataaca gctagacgta	5940
ttctgcaagg cctagggcat tatttgaagg cagaaggaaa agtggagtgc taccgaaccc	6000
tgcattcctgt tcctttgtat tcatttagtg tgaaccgtgc cttttcaagc cccaaggctg	6060
cagtggaagc ctgtaacgcc atgttgaaag agaactttcc gactgtggct tcttactgta	6120
ttattccaga gtacgatgcc tatttggaca tggttgacgg agcttcatgc tgcttagaca	6180
ctgccagttt ttgccctgca aagctgcgca gctttccaaa gaaacactcc tatttggaac	6240
ccacaatacg atcggcagtg ccttcagcga tccagaacac gctccagaac gtcctggcag	6300
ctgccacaaa aagaaattgc aatgtcacgc aaatgagaga attgcccgtg ttggattcgg	6360
cggcctttta tgtggaatgc ttcaagaaat atgcgtgtaa taatgaatat tgggaaacgt	6420
ttaaagaaaa ccccatcagg cttactgaag aaaacgtggg aaattacatt accaaattaa	6480
aaggacaaa agctgctgct ctttttgcga agacacataa tttgaatatg ttgcaggaca	6540
taccaatgga caggtttgta atggacttaa agagagacgt gaaagtgact ccaggaacaa	6600
aacatactga agaacggccc aaggtacagg tgatccaggc tgccgatccg ctagcaacag	6660
cgtatctgtg cggaatccac cgagagctgg ttaggagatt aaatgcggtc ctgcttccga	6720
acattcatac actgtttgat atgtcggctg aagactttga cgctattata gccgagcact	6780

tccagcctgg	ggattgtggt	ctggaaactg	acatcgcgtc	gtttgataaa	agtgaggacg	6840
acgccatggc	tctgaccgcg	ttaatgattc	tggaagactt	aggtgtggac	gcagagctgt	6900
tgacgctgat	tgaggcggct	ttcggcgaaa	tttcatcaat	acatttgccc	actaaaacta	6960
aatttaaatt	cggagccatg	atgaaatctg	gaatgttcct	cacactgttt	gtgaacacag	7020
tcattaacat	tgtaatcgca	agcagagtgt	tgagagaacg	gctaaccgga	tcaccatgtg	7080
cagcattcat	tggagatgac	aatatcgtga	aaggagtcaa	atcggacaaa	ttaatggcag	7140
acaggtgcg	cacctgggtg	aatatggaag	tcaagattat	agatgctgtg	gtgggcgaga	7200
aagcgcctta	tttctgtgga	gggtttat	tgtgtgactc	cgtgaccggc	acagcgtgcc	7260
gtgtggcaga	ccccctaaaa	aggctgttta	agcttggcaa	acctctggca	gcagacgatg	7320
aacatgatga	tgacaggaga	agggcattgc	atgaagagtc	aacacgctgg	aaccgagtgg	7380
gtattctttc	agagctgtgc	aaggcagtag	aatcaaggta	tgaaaccgta	ggaacttcca	7440
tcatagttat	ggccatgact	actctagcta	gcagtgttaa	atcattcagc	tacctgagag	7500
gggcccctat	aactctctac	ggctaacctg	aatggactac	gactctagaa	tagtctttaa	7560
ttaaagtccg	ccatatgaga	tggaagatgc	caaaaacatt	aagaagggcc	cagcgccatt	7620
ctaccactc	gaagacggga	ccgccggcga	gcagctgcac	aaagccatga	agcgctacgc	7680
cctggtgccc	ggcaccatcg	cctttaccga	cgcacatatc	gagggtggaca	ttacctacgc	7740
cgagtacttc	gagatgagcg	ttcggctggc	agaagctatg	aagcgctatg	ggctgaatac	7800
aaaccatcgg	atcgtggtgt	gcagcgagaa	tagcttgacg	ttcttcatgc	ccgtgttggg	7860
tgccctgttc	atcgggtgtg	ctgtggcccc	agctaacgac	atctacaacg	agcgcgagct	7920
gctgaacagc	atgggcatca	gccagcccac	cgtcgtattc	gtgagcaaga	aagggtgca	7980
aaagatcctc	aacgtgcaaa	agaagctacc	gatcatacaa	aagatcatca	tcatggatag	8040
caagaccgac	taccagggtc	tccaaagcat	gtacaccttc	gtgacttccc	atttgccacc	8100
cggcttcaac	gagtacgact	tcgtgcccga	gagcttcgac	cgggacaaaa	ccatcgccct	8160
gatcatgaac	agtagtggca	gtaccggatt	gcccgaaggc	gtagccctac	cgcaccgcac	8220
cgcttgtgtc	cgattcagtc	atgcccgcga	ccccatcttc	ggcaaccaga	tcatccccga	8280
caccgctatc	ctcagcgtgg	tgccatttca	ccacggcttc	ggcatgttca	ccacgctggg	8340
ctacttgatc	tgcggctttc	gggtcgtgct	catgtaccgc	ttcaggaggg	agctattctt	8400

gcgcagcttg caagactata agattcaatc tgccctgctg gtgcccacac tatttagctt	8460
cttcgctaag agcactctca tcgacaagta cgacctaaagc aacttgacag agatcgccag	8520
cggcggggcg ccgctcagca aggaggtagg tgaggccgtg gccaaacgct tccacctacc	8580
aggcatccgc cagggctacg gcctgacaga aacaaccagc gccattctga tcacccccga	8640
aggggacgac aagcctggcg cagtaggcaa ggtggtgccc ttcttcgagg ctaaggtggt	8700
ggacttggac accggtaaga cactgggtgt gaaccagcgc ggcgagctgt gcgtccgtgg	8760
ccccatgata atgagcggct acgttaacaa ccccgaggct acaaacgctc tcatacgaaa	8820
ggacggctgg ctgcacagcg gcgacatcgc ctactgggac gaggacgagc acttcttcat	8880
cgtggaccgg ctgaagagcc tgatcaaata caagggtac caggtagccc cagccgaact	8940
ggagagcatc ctgctgcaac accccaacat cttcgacgcc ggggtcgccg gcctgcccga	9000
cgacgatgcc ggcgagctgc ccgccgcagt cgtcgtgctg gaacacggta aaaccatgac	9060
cgagaaggag atcgtggact atgtggccag ccaggttaca accgccaaga agctgcgcgg	9120
tgggtttgtg ttcgtggacg aggtgcctaa aggactgacc ggcaagttgg acgcccga	9180
gatccgcgag attctcatta aggccaagaa gggcggcaag atcgccgtgt aattcgaacg	9240
gccgtatcac gcccaaacat ttacagccgc ggtgtcaaaa accgcgtgga cgtggttaac	9300
atccctgctg ggaggatcag ccgtaattat tataattggc ttggtgctgg ctactattgt	9360
ggccatgtac gtgctgacca accagaaaca taattgaata cagcagcaat tggcaagctg	9420
cttacataga actcgcggcg attggcatgc cgccttaaaa tttttatttt attttttctt	9480
ttcttttccg aatcggattt tgtttttaat atttcaaaaa aaaaaaaaaa aaaaaaaaaa	9540
aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaa	9595

<210> 16

<211> 139

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 16

Pro	Ser	Ser	Leu	Ser	Ala	Ser	Val	Gly	Asp	Arg	Val	Thr	Ile	Thr	Cys
1				5				10						15	

Arg Ala Ser Gln Ser Ile Asn Ser Tyr Leu Asp Trp Tyr Gln Gln Lys
20 25 30

Pro Gly Lys Ala Pro Lys Leu Leu Ile Tyr Ala Ala Ser Ser Leu Gln
35 40 45

Ser Gly Val Pro Ser Arg Phe Ser Gly Ser Gly Ser Gly Thr Asp Phe
50 55 60

Thr Leu Thr Ile Ser Ser Leu Gln Pro Glu Asp Phe Ala Thr Tyr Tyr
65 70 75 80

Cys Gln Gln Tyr Tyr Ser Thr Pro Phe Thr Phe Gly Pro Gly Thr Lys
85 90 95

Val Glu Ile Lys Arg Thr Val Ala Ala Pro Ser Val Phe Ile Phe Pro
100 105 110

Pro Ser Asp Glu Gln Leu Lys Ser Gly Thr Ala Ser Val Val Cys Leu
115 120 125

Leu Asn Asn Phe Tyr Pro Arg Glu Ala Lys Val
130 135

<210> 17

<211> 167

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic
polypeptide

<400> 17

Gly Val Val Gln Pro Gly Arg Ser Leu Arg Leu Ser Cys Ala Ala Ser
1 5 10 15

Gly Phe Thr Phe Ser Ser Tyr Gly Met His Trp Val Arg Gln Ala Pro
20 25 30

Gly Lys Gly Leu Glu Trp Val Ala Val Ile Trp Tyr Asp Gly Ser Asn
35 40 45

Lys Tyr Tyr Ala Asp Ser Val Lys Gly Arg Phe Thr Ile Ser Arg Asp
50 55 60

Asn Ser Lys Asn Thr Leu Tyr Leu Gln Met Asn Ser Leu Arg Ala Glu
65 70 75 80

Asp Thr Ala Val Tyr Tyr Cys Ala Arg Asp Pro Arg Gly Ala Thr Leu
85 90 95

Tyr Tyr Tyr Tyr Tyr Gly Met Asp Val Trp Gly Gln Gly Thr Thr Val
100 105 110

Thr Val Ser Ser Ala Ser Thr Lys Gly Pro Ser Val Phe Pro Leu Ala
115 120 125

Pro Cys Ser Arg Ser Thr Ser Glu Ser Thr Ala Ala Leu Gly Cys Leu
130 135 140

Val Lys Asp Tyr Phe Pro Glu Pro Val Thr Val Ser Trp Asn Ser Gly
145 150 155 160

Ala Leu Thr Ser Gly Val His
165

<210> 18

<211> 10

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic
peptide

<400> 18

Gly Phe Thr Phe Ser Ser Tyr Gly Met His
1 5 10

<210> 19

<211> 15

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic

peptide

<400> 19

Val Ile Trp Tyr Asp Gly Ser Asn Lys Tyr Tyr Ala Asp Ser Val
1 5 10 15

<210> 20

<211> 16

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic
peptide

<400> 20

Asp Pro Arg Gly Ala Thr Leu Tyr Tyr Tyr Tyr Tyr Gly Met Asp Val
1 5 10 15

<210> 21

<211> 11

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic
peptide

<400> 21

Arg Ala Ser Gln Ser Ile Asn Ser Tyr Leu Asp
1 5 10

<210> 22

<211> 7

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic
peptide

<400> 22

Ala Ala Ser Ser Leu Gln Ser
1 5

<210> 23

<211> 9

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic peptide

<400> 23

Gln Gln Tyr Tyr Ser Thr Pro Phe Thr
1 5

<210> 24

<211> 108

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 24

Glu Ile Val Leu Thr Gln Ser Pro Gly Thr Leu Ser Leu Ser Pro Gly
1 5 10 15

Glu Arg Ala Thr Leu Ser Cys Arg Ala Ser Gln Arg Val Ser Ser Ser
20 25 30

Tyr Leu Ala Trp Tyr Gln Gln Lys Pro Gly Gln Ala Pro Arg Leu Leu
35 40 45

Ile Tyr Asp Ala Ser Ser Arg Ala Thr Gly Ile Pro Asp Arg Phe Ser
50 55 60

Gly Ser Gly Ser Gly Thr Asp Phe Thr Leu Thr Ile Ser Arg Leu Glu
65 70 75 80

Pro Glu Asp Phe Ala Val Tyr Tyr Cys Gln Gln Tyr Gly Ser Leu Pro
85 90 95

Trp Thr Phe Gly Gln Gly Thr Lys Val Glu Ile Lys
100 105

<210> 25

<211> 121

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic

polypeptide

<400> 25

Glu Val Gln Leu Val Glu Ser Gly Gly Gly Leu Val Gln Pro Gly Gly
1 5 10 15

Ser Leu Arg Leu Ser Cys Ala Ala Ser Gly Phe Thr Phe Ser Arg Tyr
20 25 30

Trp Met Ser Trp Val Arg Gln Ala Pro Gly Lys Gly Leu Glu Trp Val
35 40 45

Ala Asn Ile Lys Gln Asp Gly Ser Glu Lys Tyr Tyr Val Asp Ser Val
50 55 60

Lys Gly Arg Phe Thr Ile Ser Arg Asp Asn Ala Lys Asn Ser Leu Tyr
65 70 75 80

Leu Gln Met Asn Ser Leu Arg Ala Glu Asp Thr Ala Val Tyr Tyr Cys
85 90 95

Ala Arg Glu Gly Gly Trp Phe Gly Glu Leu Ala Phe Asp Tyr Trp Gly
100 105 110

Gln Gly Thr Leu Val Thr Val Ser Ser
115 120

<210> 26

<211> 5

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic
peptide

<400> 26

Arg Tyr Trp Met Ser
1 5

<210> 27

<211> 17

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic peptide

<400> 27

Asn Ile Lys Gln Asp Gly Ser Glu Lys Tyr Tyr Val Asp Ser Val Lys
1 5 10 15

Gly

<210> 28

<211> 12

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic peptide

<400> 28

Glu Gly Gly Trp Phe Gly Glu Leu Ala Phe Asp Tyr
1 5 10

<210> 29

<211> 12

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic peptide

<400> 29

Arg Ala Ser Gln Arg Val Ser Ser Ser Tyr Leu Ala
1 5 10

<210> 30

<211> 7

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic peptide

<400> 30

Asp Ala Ser Ser Arg Ala Thr
1 5

<210> 31
<211> 9
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic peptide

<400> 31
Gln Gln Tyr Gly Ser Leu Pro Trp Thr
1 5

<210> 32
<211> 2019
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 32
gccccgggcat ttaa atgcga tcgcatcgat tacgactcta gaatagtcta gtccgcaggc 60
caccatgcag atcttcgtga agaccctgac cggcaagacc atcaccctag aggtggagcc 120
cagtgcacc atcgagaacg tgaaggccaa gatccaggat aaagagggca tccccctga 180
ccagcagagg ctgatctttg ccggcaagca gctggaagat ggccgcaccc tctctgatta 240
caacatccag aaggagtcaa ccctgcacct ggtccttcgc ctgagagggtg ccatgtttca 300
ggcgctgagc gaaggctgca ccccgatatga tattaaccag atgctgaacg tgctgggcga 360
tcatcaggctc tcaggccttg agcagcttga gagtataatc aactttgaaa aactgactga 420
atggaccagt tcta atgtta tgcctatcct gtctcctctg acaaagggca tcctgggctt 480
cgtgtttacc ctgaccgtgc cttctgagag aggacttagc tgcattagcg aagcggatgc 540
gaccaccccg gaaagcgga acctgggcga agaaattctg agccagctgt atctttggcc 600
aagggtgacc taccattccc ctagttatgc ttaccaccaa tttgaaagac gagccaaata 660
taaaagacac ttccccggct ttggccagag cctgctgttt ggctaccctg tgtacgtgtt 720
cggcgattgc gtgcagggcg attgggatgc gattcgcttt cgctattgcg cgccgccggg 780
ctatgcgctg ctgcgctgca acgataccaa ctatagcgct ctgctggctg tgggggcccct 840
agaaggaccc aggaatcagg actggcttgg tgtcccaaga caacttgtaa ctcggatgca 900

ggctattcag aatgccggcc tgtgtaccct ggtggccatg ctggaagaga caatcttctg	960
gctgcaagcg tttctgatgg cgctgaccga tagcggcccg aaaaccaaca ttattgtgga	1020
tagccagtat gtgatgggca ttagcaaacc gagctttcag gaatttgtgg attgggaaaa	1080
cgtgagcccg gaactgaaca gcaccgatca gccgttttgg caagccggaa tcctggccag	1140
aaatctggtg cctatggtgg ccacagtgcg gggccagaac ctgaagtacc agggtcagtc	1200
actagtcatc tctgcttcta tcattgtctt caacctgctg gaactggaag gtgattatcg	1260
agatgatggc aacgtgtggg tgcatacccc gctgagcccg cgcaccctga acgcgtgggt	1320
gaaagcggcg gaagaaaaaa aaggtattcc agttcaccta gagctggcca gtatgaccaa	1380
catggagctc atgagcagta ttgtgcatca gcaggtcaga acatacggcc ccgtgttcat	1440
gtgtctcggc ggactgctta caatggtggc tgggtgctgtg tggctgacag tgcgagtgtc	1500
cgagctgttc cgggccgcgc agctggccaa cgacgtgggc ctccagatca tggagctttg	1560
tgggtgcagcg tttcgccagg tgtgccatac caccgtgccg tggccgaacg cgagcctgac	1620
cccgaaatgg aacaacgaaa ccaccagcc ccagatcgcc aactgcagcg tgtatgactt	1680
ttttgtgtgg ctccattatt attctgttcg agacacactt tggccaaggg tgacctacca	1740
tatgaacaaa tatgcgtatc atatgctgga aagacgagcc aaatataaaa gaggaccagg	1800
acctggcgct aaatttgtgg ccgcctggac actgaaagcc gctgctgggc ctggacctgg	1860
ccagtacatc aaggccaaca gcaagttcat cggcatcacc gaactcggac ccggaccagg	1920
ctgatgattt cgaaatttaa ataagcttgc ggccgctagg gataacaggg taattatcac	1980
gcccacacat ttacagccgc ggtgtcaaaa accgcgtgg	2019

<210> 33

<211> 619

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 33

Met	Gln	Ile	Phe	Val	Lys	Thr	Leu	Thr	Gly	Lys	Thr	Ile	Thr	Leu	Glu
1				5					10					15	

Val Glu Pro Ser Asp Thr Ile Glu Asn Val Lys Ala Lys Ile Gln Asp
20 25 30

Lys Glu Gly Ile Pro Pro Asp Gln Gln Arg Leu Ile Phe Ala Gly Lys
35 40 45

Gln Leu Glu Asp Gly Arg Thr Leu Ser Asp Tyr Asn Ile Gln Lys Glu
50 55 60

Ser Thr Leu His Leu Val Leu Arg Leu Arg Gly Ala Met Phe Gln Ala
65 70 75 80

Leu Ser Glu Gly Cys Thr Pro Tyr Asp Ile Asn Gln Met Leu Asn Val
85 90 95

Leu Gly Asp His Gln Val Ser Gly Leu Glu Gln Leu Glu Ser Ile Ile
100 105 110

Asn Phe Glu Lys Leu Thr Glu Trp Thr Ser Ser Asn Val Met Pro Ile
115 120 125

Leu Ser Pro Leu Thr Lys Gly Ile Leu Gly Phe Val Phe Thr Leu Thr
130 135 140

Val Pro Ser Glu Arg Gly Leu Ser Cys Ile Ser Glu Ala Asp Ala Thr
145 150 155 160

Thr Pro Glu Ser Ala Asn Leu Gly Glu Glu Ile Leu Ser Gln Leu Tyr
165 170 175

Leu Trp Pro Arg Val Thr Tyr His Ser Pro Ser Tyr Ala Tyr His Gln
180 185 190

Phe Glu Arg Arg Ala Lys Tyr Lys Arg His Phe Pro Gly Phe Gly Gln
195 200 205

Ser Leu Leu Phe Gly Tyr Pro Val Tyr Val Phe Gly Asp Cys Val Gln
210 215 220

Gly Asp Trp Asp Ala Ile Arg Phe Arg Tyr Cys Ala Pro Pro Gly Tyr
225 230 235 240

Ala Leu Leu Arg Cys Asn Asp Thr Asn Tyr Ser Ala Leu Leu Ala Val
245 250 255

Gly Ala Leu Glu Gly Pro Arg Asn Gln Asp Trp Leu Gly Val Pro Arg
260 265 270

Gln Leu Val Thr Arg Met Gln Ala Ile Gln Asn Ala Gly Leu Cys Thr
275 280 285

Leu Val Ala Met Leu Glu Glu Thr Ile Phe Trp Leu Gln Ala Phe Leu
290 295 300

Met Ala Leu Thr Asp Ser Gly Pro Lys Thr Asn Ile Ile Val Asp Ser
305 310 315 320

Gln Tyr Val Met Gly Ile Ser Lys Pro Ser Phe Gln Glu Phe Val Asp
325 330 335

Trp Glu Asn Val Ser Pro Glu Leu Asn Ser Thr Asp Gln Pro Phe Trp
340 345 350

Gln Ala Gly Ile Leu Ala Arg Asn Leu Val Pro Met Val Ala Thr Val
355 360 365

Gln Gly Gln Asn Leu Lys Tyr Gln Gly Gln Ser Leu Val Ile Ser Ala
370 375 380

Ser Ile Ile Val Phe Asn Leu Leu Glu Leu Glu Gly Asp Tyr Arg Asp
385 390 395 400

Asp Gly Asn Val Trp Val His Thr Pro Leu Ser Pro Arg Thr Leu Asn
405 410 415

Ala Trp Val Lys Ala Val Glu Glu Lys Lys Gly Ile Pro Val His Leu
420 425 430

Glu Leu Ala Ser Met Thr Asn Met Glu Leu Met Ser Ser Ile Val His
435 440 445

Gln Gln Val Arg Thr Tyr Gly Pro Val Phe Met Cys Leu Gly Gly Leu
450 455 460

Leu Thr Met Val Ala Gly Ala Val Trp Leu Thr Val Arg Val Leu Glu
465 470 475 480

Leu Phe Arg Ala Ala Gln Leu Ala Asn Asp Val Val Leu Gln Ile Met
485 490 495

Glu Leu Cys Gly Ala Ala Phe Arg Gln Val Cys His Thr Thr Val Pro
500 505 510

Trp Pro Asn Ala Ser Leu Thr Pro Lys Trp Asn Asn Glu Thr Thr Gln
515 520 525

Pro Gln Ile Ala Asn Cys Ser Val Tyr Asp Phe Phe Val Trp Leu His
530 535 540

Tyr Tyr Ser Val Arg Asp Thr Leu Trp Pro Arg Val Thr Tyr His Met
545 550 555 560

Asn Lys Tyr Ala Tyr His Met Leu Glu Arg Arg Ala Lys Tyr Lys Arg
565 570 575

Gly Pro Gly Pro Gly Ala Lys Phe Val Ala Ala Trp Thr Leu Lys Ala
580 585 590

Ala Ala Gly Pro Gly Pro Gly Gln Tyr Ile Lys Ala Asn Ser Lys Phe
595 600 605

Ile Gly Ile Thr Glu Leu Gly Pro Gly Pro Gly
610 615

<210> 34

<211> 1638

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic
polynucleotide

<400> 34

atggccggga tgttccaggc actgtccgaa ggctgcacac cctatgatat taaccagatg	60
ctgaatgtcc tgggagacca ccaggctctt ggcctggagc agctggagag catcatcaac	120
ttcgagaagc tgaccgagtg gacaagctcc aatgtgatgc ctatcctgtc cccactgacc	180
aagggcattc tgggcttcgt gtttaccctg acagtgccct ctgagcgggg cctgtcttgc	240
atcagcgagg cagacgcaac cacaccagag tccgccaatc tgggcgagga gatcctgtct	300
cagctgtacc tgtggccccg ggtgacatat cactccccct cttacgccta tcaccagttc	360
gagcggagag ccaagtacaa gagacacttc ccaggctttg gccagtctct gctgttcggc	420
taccccgtgt acgtgttcgg cgattgcgtg cagggcgact gggatgccat ccggtttaga	480
tactgcgcac cacctggata tgcactgctg aggtgtaacg acaccaatta ttccgccctg	540
ctggcagtgg gcgccctgga gggccctcgc aatcaggatt ggctgggcgt gccaaggcag	600
ctggtgacac gcatgcaggc catccagaac gcaggcctgt gcaccctggg ggcaatgctg	660
gaggagacaa tcttctggct gcaggccttt ctgatggccc tgaccgacag cggccccaag	720
acaaacatca tcgtggattc ccagtacgtg atgggcatct ccaagccttc tttccaggag	780
tttgtggact gggagaacgt gagcccagag ctgaattcca ccgatcagcc attctggcag	840
gcaggaatcc tggcaaggaa cctggtgcct atggtggcca cagtgcaggg ccagaatctg	900
aagtaccagg gccagagcct ggtcatcagc gcctccatca tcgtgtttaa cctgctggag	960
ctggagggcg actatcgga cgatggcaac gtgtgggtgc acacccact gagccccaga	1020
acactgaacg cctgggtgaa ggccgtggag gagaagaagg gcatcccagt gcacctggag	1080
ctggcctcca tgaccaatat ggagctgatg tctagcatcg tgcaccagca ggtgaggaca	1140
tacggacccg tgttcatgtg cctgggaggc ctgctgacca tgggtggcagg agccgtgtgg	1200
ctgacagtgc gggtgctgga gctgttcaga gccgcccagc tggccaacga tgtggtgctg	1260
cagatcatgg agctgtgcgg agcagccttt cgccagggtgt gccacaccac agtgccatgg	1320
cccaatgcct ccctgacccc caagtggaac aatgagacaa cacagcctca gatcgccaac	1380
tgtagcgtgt acgacttctt cgtgtggctg cactactata gcgtgaggga taccctgtgg	1440
ccccgcgtga cataccacat gaataagtac gcctatcaca tgctggagag gcgcgccaag	1500
tataagagag gccctggccc aggcgcaaag tttgtggcag catggaccct gaaggccgcc	1560
gccggccccg gccccggcca gtatatcaag gctaacagta agttcattgg aatcacagag	1620

ctgggacccg gacctgga

1638

<210> 35

<211> 546

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic
polypeptide

<400> 35

Met	Ala	Gly	Met	Phe	Gln	Ala	Leu	Ser	Glu	Gly	Cys	Thr	Pro	Tyr	Asp
1				5					10					15	

Ile	Asn	Gln	Met	Leu	Asn	Val	Leu	Gly	Asp	His	Gln	Val	Ser	Gly	Leu
			20					25					30		

Glu	Gln	Leu	Glu	Ser	Ile	Ile	Asn	Phe	Glu	Lys	Leu	Thr	Glu	Trp	Thr
		35					40					45			

Ser	Ser	Asn	Val	Met	Pro	Ile	Leu	Ser	Pro	Leu	Thr	Lys	Gly	Ile	Leu
	50					55					60				

Gly	Phe	Val	Phe	Thr	Leu	Thr	Val	Pro	Ser	Glu	Arg	Gly	Leu	Ser	Cys
65					70					75					80

Ile	Ser	Glu	Ala	Asp	Ala	Thr	Thr	Pro	Glu	Ser	Ala	Asn	Leu	Gly	Glu
				85					90					95	

Glu	Ile	Leu	Ser	Gln	Leu	Tyr	Leu	Trp	Pro	Arg	Val	Thr	Tyr	His	Ser
			100					105					110		

Pro	Ser	Tyr	Ala	Tyr	His	Gln	Phe	Glu	Arg	Arg	Ala	Lys	Tyr	Lys	Arg
		115					120					125			

His	Phe	Pro	Gly	Phe	Gly	Gln	Ser	Leu	Leu	Phe	Gly	Tyr	Pro	Val	Tyr
	130					135					140				

Val	Phe	Gly	Asp	Cys	Val	Gln	Gly	Asp	Trp	Asp	Ala	Ile	Arg	Phe	Arg
145					150					155					160

Tyr Cys Ala Pro Pro Gly Tyr Ala Leu Leu Arg Cys Asn Asp Thr Asn

165

170

175

Tyr Ser Ala Leu Leu Ala Val Gly Ala Leu Glu Gly Pro Arg Asn Gln
 180 185 190

Asp Trp Leu Gly Val Pro Arg Gln Leu Val Thr Arg Met Gln Ala Ile
 195 200 205

Gln Asn Ala Gly Leu Cys Thr Leu Val Ala Met Leu Glu Glu Thr Ile
 210 215 220

Phe Trp Leu Gln Ala Phe Leu Met Ala Leu Thr Asp Ser Gly Pro Lys
 225 230 235 240

Thr Asn Ile Ile Val Asp Ser Gln Tyr Val Met Gly Ile Ser Lys Pro
 245 250 255

Ser Phe Gln Glu Phe Val Asp Trp Glu Asn Val Ser Pro Glu Leu Asn
 260 265 270

Ser Thr Asp Gln Pro Phe Trp Gln Ala Gly Ile Leu Ala Arg Asn Leu
 275 280 285

Val Pro Met Val Ala Thr Val Gln Gly Gln Asn Leu Lys Tyr Gln Gly
 290 295 300

Gln Ser Leu Val Ile Ser Ala Ser Ile Ile Val Phe Asn Leu Leu Glu
 305 310 315 320

Leu Glu Gly Asp Tyr Arg Asp Asp Gly Asn Val Trp Val His Thr Pro
 325 330 335

Leu Ser Pro Arg Thr Leu Asn Ala Trp Val Lys Ala Val Glu Glu Lys
 340 345 350

Lys Gly Ile Pro Val His Leu Glu Leu Ala Ser Met Thr Asn Met Glu
 355 360 365

Leu Met Ser Ser Ile Val His Gln Gln Val Arg Thr Tyr Gly Pro Val
 370 375 380

Phe Met Cys Leu Gly Gly Leu Leu Thr Met Val Ala Gly Ala Val Trp
385 390 395 400

Leu Thr Val Arg Val Leu Glu Leu Phe Arg Ala Ala Gln Leu Ala Asn
405 410 415

Asp Val Val Leu Gln Ile Met Glu Leu Cys Gly Ala Ala Phe Arg Gln
420 425 430

Val Cys His Thr Thr Val Pro Trp Pro Asn Ala Ser Leu Thr Pro Lys
435 440 445

Trp Asn Asn Glu Thr Thr Gln Pro Gln Ile Ala Asn Cys Ser Val Tyr
450 455 460

Asp Phe Phe Val Trp Leu His Tyr Tyr Ser Val Arg Asp Thr Leu Trp
465 470 475 480

Pro Arg Val Thr Tyr His Met Asn Lys Tyr Ala Tyr His Met Leu Glu
485 490 495

Arg Arg Ala Lys Tyr Lys Arg Gly Pro Gly Pro Gly Ala Lys Phe Val
500 505 510

Ala Ala Trp Thr Leu Lys Ala Ala Ala Gly Pro Gly Pro Gly Gln Tyr
515 520 525

Ile Lys Ala Asn Ser Lys Phe Ile Gly Ile Thr Glu Leu Gly Pro Gly
530 535 540

Pro Gly
545

<210> 36

<211> 2019

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic
polynucleotide

<400> 36

gccccgggcat	ttaa	atgcga	tcgcatcgat	tacgactcta	gaatagtcta	gtccgcaggc	60
caccatgcag	atcttcgtga	agaccctgac	cggcaagacc	atcacccctag	aggtggagcc		120
cagtgcacc	atcgagaacg	tgaaggccaa	gatccaggat	aaagagggca	tccccctga		180
ccagcagagg	ctgatctttg	ccggcaagca	gctggaagat	ggccgcaccc	tctctgatta		240
caacatccag	aaggagtcaa	ccctgcacct	ggtccttcgc	ctgagagggtg	ccatgtttca		300
ggcgctgagc	gaaggctgca	ccccgtatga	tattaaccag	atgctgaacg	tgctgggcga		360
tcatcagttt	aagcacatca	aagcctttga	ccggacattt	gctaacaacc	caggtcccat		420
ggttggtgtt	gccacacctg	ggcctatcct	gtctcctctg	acaaagggca	tcctgggctt		480
cgtgtttacc	ctgaccgtgc	cttctgagag	aggacttagc	tgcattagcg	aagcggatgc		540
gaccaccccg	gaaagcgga	acctgggcga	agaaattctg	agccagctgt	atctttggcc		600
aagggtgacc	taccattccc	ctagttatgc	ttaccaccaa	tttgaaagac	gagccaaata		660
taaaagacac	ttccccggct	ttggccagag	cctgctgttt	ggctaccctg	tgtacgtgtt		720
cggcgattgc	gtgcagggcg	attgggatgc	gattcgcttt	cgctattgcg	cgccgccggg		780
ctatgcgctg	ctgcgctgca	acgataccaa	ctatagcgct	ctgctggctg	tggggggcct		840
agaaggaccc	aggaatcagg	actggcttgg	tgtcccaaga	caacttgtaa	ctcggatgca		900
ggctattcag	aatgccggcc	tgtgtaccct	ggtggccatg	ctggaagaga	caatcttctg		960
gctgcaagcg	tttctgatgg	cgctgaccga	tagcggcccg	aaaaccaaca	ttattgtgga		1020
tagccagtat	gtgatgggca	ttagcaaacc	gagctttcag	gaatttgtgg	attgggaaaa		1080
cgtgagcccg	gaactgaaca	gcaccgatca	gccgttttgg	caagccggaa	tcctggccag		1140
aaatctggtg	cctatggtgg	ccacagtgca	gggccagAAC	ctgaagtacc	agggtcagtc		1200
actagtcatc	tctgcttcta	tcattgtctt	caacctgctg	gaactggaag	gtgattatcg		1260
agatgatggc	aacgtgtggg	tgcatacccc	gctgagcccg	cgcaccctga	acgcgtgggt		1320
gaaagcggtg	gaagaaaaaa	aaggtattcc	agttcaccta	gagctggcca	gtatgaccaa		1380
catggagctc	atgagcagta	ttgtgcatca	gcaggtcaga	acatacggcc	ccgtgttcac		1440
gtgtctcggc	ggactgctta	caatggtggc	tggtgctgtg	tggctgacag	tgcgagtgtc		1500
cgagctgttc	cgggccgcgc	agctggccaa	cgacgtggtc	ctccagatca	tggagctttg		1560
tgggtgcagcg	tttcgccagg	tgtgccatac	caccgtgccg	tggccgaacg	cgagcctgac		1620

```

cccgaaatgg aacaacgaaa ccaccagcc ccagatcgcc aactgcagcg tgtatgactt 1680
ttttgtgtgg ctccattatt attctgttcg agacacactt tggccaaggg tgacctacca 1740
tatgaacaaa tatgcgtatc atatgctgga aagacgagcc aaatataaaa gaggaccagg 1800
acctggcgct aaatttgtgg cgcctggac actgaaagcc gctgctggtc ctggacctgg 1860
ccagtacatc aaggccaaca gcaagttcat cggcatcacc gaactcggac ccggaccagg 1920
ctgatgattt cgaaatttaa ataagcttgc ggccgctagg gataacaggg taattatcac 1980
gcccaaacat ttacagccgc ggtgtcaaaa accgcgtgg 2019

```

<210> 37

<211> 619

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic
polypeptide

<400> 37

```

Met Gln Ile Phe Val Lys Thr Leu Thr Gly Lys Thr Ile Thr Leu Glu
1           5           10           15

```

```

Val Glu Pro Ser Asp Thr Ile Glu Asn Val Lys Ala Lys Ile Gln Asp
          20           25           30

```

```

Lys Glu Gly Ile Pro Pro Asp Gln Gln Arg Leu Ile Phe Ala Gly Lys
          35           40           45

```

```

Gln Leu Glu Asp Gly Arg Thr Leu Ser Asp Tyr Asn Ile Gln Lys Glu
          50           55           60

```

```

Ser Thr Leu His Leu Val Leu Arg Leu Arg Gly Ala Met Phe Gln Ala
65           70           75           80

```

```

Leu Ser Glu Gly Cys Thr Pro Tyr Asp Ile Asn Gln Met Leu Asn Val
          85           90           95

```

```

Leu Gly Asp His Gln Phe Lys His Ile Lys Ala Phe Asp Arg Thr Phe
          100          105          110

```

Ala Asn Asn Pro Gly Pro Met Val Val Phe Ala Thr Pro Gly Pro Ile
115 120 125

Leu Ser Pro Leu Thr Lys Gly Ile Leu Gly Phe Val Phe Thr Leu Thr
130 135 140

Val Pro Ser Glu Arg Gly Leu Ser Cys Ile Ser Glu Ala Asp Ala Thr
145 150 155 160

Thr Pro Glu Ser Ala Asn Leu Gly Glu Glu Ile Leu Ser Gln Leu Tyr
165 170 175

Leu Trp Pro Arg Val Thr Tyr His Ser Pro Ser Tyr Ala Tyr His Gln
180 185 190

Phe Glu Arg Arg Ala Lys Tyr Lys Arg His Phe Pro Gly Phe Gly Gln
195 200 205

Ser Leu Leu Phe Gly Tyr Pro Val Tyr Val Phe Gly Asp Cys Val Gln
210 215 220

Gly Asp Trp Asp Ala Ile Arg Phe Arg Tyr Cys Ala Pro Pro Gly Tyr
225 230 235 240

Ala Leu Leu Arg Cys Asn Asp Thr Asn Tyr Ser Ala Leu Leu Ala Val
245 250 255

Gly Ala Leu Glu Gly Pro Arg Asn Gln Asp Trp Leu Gly Val Pro Arg
260 265 270

Gln Leu Val Thr Arg Met Gln Ala Ile Gln Asn Ala Gly Leu Cys Thr
275 280 285

Leu Val Ala Met Leu Glu Glu Thr Ile Phe Trp Leu Gln Ala Phe Leu
290 295 300

Met Ala Leu Thr Asp Ser Gly Pro Lys Thr Asn Ile Ile Val Asp Ser
305 310 315 320

Gln Tyr Val Met Gly Ile Ser Lys Pro Ser Phe Gln Glu Phe Val Asp
325 330 335

Trp Glu Asn Val Ser Pro Glu Leu Asn Ser Thr Asp Gln Pro Phe Trp
340 345 350

Gln Ala Gly Ile Leu Ala Arg Asn Leu Val Pro Met Val Ala Thr Val
355 360 365

Gln Gly Gln Asn Leu Lys Tyr Gln Gly Gln Ser Leu Val Ile Ser Ala
370 375 380

Ser Ile Ile Val Phe Asn Leu Leu Glu Leu Glu Gly Asp Tyr Arg Asp
385 390 395 400

Asp Gly Asn Val Trp Val His Thr Pro Leu Ser Pro Arg Thr Leu Asn
405 410 415

Ala Trp Val Lys Ala Val Glu Glu Lys Lys Gly Ile Pro Val His Leu
420 425 430

Glu Leu Ala Ser Met Thr Asn Met Glu Leu Met Ser Ser Ile Val His
435 440 445

Gln Gln Val Arg Thr Tyr Gly Pro Val Phe Met Cys Leu Gly Gly Leu
450 455 460

Leu Thr Met Val Ala Gly Ala Val Trp Leu Thr Val Arg Val Leu Glu
465 470 475 480

Leu Phe Arg Ala Ala Gln Leu Ala Asn Asp Val Val Leu Gln Ile Met
485 490 495

Glu Leu Cys Gly Ala Ala Phe Arg Gln Val Cys His Thr Thr Val Pro
500 505 510

Trp Pro Asn Ala Ser Leu Thr Pro Lys Trp Asn Asn Glu Thr Thr Gln
515 520 525

Pro Gln Ile Ala Asn Cys Ser Val Tyr Asp Phe Phe Val Trp Leu His
530 535 540

Tyr Tyr Ser Val Arg Asp Thr Leu Trp Pro Arg Val Thr Tyr His Met
545 550 555 560

Asn Lys Tyr Ala Tyr His Met Leu Glu Arg Arg Ala Lys Tyr Lys Arg
565 570 575

Gly Pro Gly Pro Gly Ala Lys Phe Val Ala Ala Trp Thr Leu Lys Ala
580 585 590

Ala Ala Gly Pro Gly Pro Gly Gln Tyr Ile Lys Ala Asn Ser Lys Phe
595 600 605

Ile Gly Ile Thr Glu Leu Gly Pro Gly Pro Gly
610 615

<210> 38
<211> 228
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
polynucleotide

<400> 38
atgcagatct tcgtgaagac cctgaccggc aagaccatca ccctagaggt ggagcccagt 60
gacaccatcg agaacgtgaa ggccaagatc caggataaag agggcatccc ccctgaccag 120
cagaggctga tctttgccgg caagcagctg gaagatggcc gcaccctctc tgattacaac 180
atccagaagg agtcaaccct gcacctggtc cttcgctga gaggtggc 228

<210> 39
<211> 228
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
polynucleotide

<400> 39
atgcagatct tcgtgaagac cctgaccggc aagaccatca ccctagaggt ggagcccagt 60
gacaccatcg agaacgtgaa ggccaagatc caggataaag agggcatccc ccctgaccag 120
cagaggctga tctttgccgg caagcagctg gaagatggcc gcaccctctc tgattacaac 180

atccagaagg agtcaaccct gcacctggtc cttcgctga gaggtgcc 228

<210> 40

<211> 78

<212> DNA

<213> Homo sapiens

<400> 40

atggccgtca tggcgcccg aaccctcgtc ctgctactct cgggggctct ggccctgacc 60

cagacctggg cgggctct 78

<210> 41

<211> 201

<212> DNA

<213> Homo sapiens

<400> 41

ccgtcttccc agcccacat cccatcgctg ggcatcattg ctggcctggg tctctttgga 60

gctgtgatca ctggagctgt ggtcgctgct gtgatgtgga ggaggaagag ctcatataga 120

aaaggaggga gctactctca ggctgcaagc agtgacagtg ccaggggctc tgatgtgtct 180

ctcacagctt gtaaagtgtg a 201

<210> 42

<211> 60

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 42

atggagaccg atacactgct gctgtgggtg ctgctcctgt gggtgccagg aagcacaggc 60

<210> 43

<211> 3178

<212> DNA

<213> Homo sapiens

<400> 43

ggcaccgatt cggggcctgc ccggacttcg ccgcacgctg cagaacctcg ccagcgccc 60

accatgcccc ggcagctcag cgcggcggcc gcgctcttcg cgtccctggc cgtaattttg 120

cacgatggca gtcaaagag agcaaaagca tttccagaaa ccagagatta ttctcaacct 180

actgcagcag caacagtaca ggacataaaa aaacctgtcc agcaaccagc taagcaagca	240
cctcaccaaa ctttagcagc aagattcatg gatggtcata tcacctttca aacagcggcc	300
acagtaaaaa ttccaacaac taccacagca actacaaaaa aacttgcaac caccagccca	360
attacctaca ccctggtcac aaccagggcc acaccaaca actcacacac agtcctcca	420
gttactgaag ttacagtcgg ccctagctta gccccttatt cactgccacc caccatcacc	480
ccaccagctc atacagctgg aaccagttca tcaaccgtca gccacacaac tgggaacacc	540
actcaacca gtaaccagac cacccttcca gcaactttat cgatagcact gcacaaaagc	600
acaaccggtc agaagcctga tcaaccacc catgccccag gaacaacggc agctgcccac	660
aataccacc gcacagctgc acctgcctcc acggttcctg ggcccaccct tgcacctcag	720
ccatcgtcag tcaagactgg aatttatcag gttctaaacg gaagcagact ctgtataaaa	780
gcagagatgg ggatacagct gattgttcaa gacaaggagt cggttttttc acctcggaga	840
tacttcaaca tcgaccccaa cgcaacgcaa gcctctggga actgtggcac ccgaaaatcc	900
aaccttctgt tgaattttca gggcggattt gtgaatctca catttaccaa ggatgaagaa	960
tcatattata tcagtgaagt gggagcctat ttgaccgtct cagatccaga gacagtttac	1020
caaggaatca aacatgcggt ggtgatgttc cagacagcag tcgggcattc cttcaagtgc	1080
gtgagtgaac agagcctcca gttgtcagcc cacctgcagg tgaaaacaac cgatgtccaa	1140
cttcaagcct ttgattttga agatgaccac tttggaaatg tggatgagtg ctctctgac	1200
tacacaattg tgcttctgt gattggggcc atcgtggttg gtctctgcct tatgggtatg	1260
ggtgtctata aaatccgcct aaggtgtcaa tcatctggat accagagaat ctaattgttg	1320
cccgggggga atgaaaataa tggaatttag agaactcttt catcccttcc aggatggatg	1380
ttgggaaatt ccctcagagt gtgggtcctt caaacaatgt aaaccaccat cttctattca	1440
aatgaagtga gtcattgttg atttaagttc aggcagcaca tcaatttcta aatacttttt	1500
gtttatttta tgaaagatat agtgagctgt ttattttcta gtttccttta gaatatttta	1560
gccactcaaa gtcaacattt gagatatgtt gaattaacat aatatatgta aagtagaata	1620
agccttcaaa ttataaacca aggggtcaatt gtaactaata ctactgtgtg tgcattgaag	1680
attttatttt acccttgatc ttaacaaagc ctttgctttg ttatcaaag gactttcagt	1740
gcttttacta tctgtgtttt atggtttcat gtaacataca tattcctggg tagcactta	1800

actccttttc cactttaaat ttgtttttgt tttttgagac ggagtttcac tcttgtcacc	1860
caggctggag tacagtggca cgatctcggc ttatggcaac ctccgcctcc cgggttcaag	1920
tgattctcct gcttcagctt cccgagtagc tgggattaca ggcacacact accacgcctg	1980
gctaattttt gtatttttat tatagacggg tttcaccatg ttggccagac tggctttgaa	2040
ctcttgacct caggtgatcc acccacctca gcctcccaaa gtgctgggat tacaggcatg	2100
agccattgcg cccggcctta aatgtttttt ttaatcatca aaaagaacaa catatctcag	2160
gttgtctaag tgtttttatg taaaaccaac aaaaagaaca aatcagctta tattttttat	2220
cttgatgact cctgctccag aattgctaga ctaagaatta ggtggctaca gatggtagaa	2280
ctaaacaata agcaagagac aataataatg gcccttaatt attaacaaag tgccagagtc	2340
taggctaagc actttatcta tatctcattt cattctcaca acttataagt gaatgagtaa	2400
actgagactt aagggaactg aatcacttaa atgtcacctg gctaactgat ggagagcca	2460
gagcttgaat tcatgttggg ctgacatcaa ggtctttggg cttctcccta caccaagtta	2520
cctacaagaa caatgacacc acactctgcc tgaaggctca cacctcatac cagcatacgc	2580
tcaccttaca gggaaatggg tttatccagg atcatgagac attagggtag atgaaaggag	2640
agctttgcag ataacaaaat agcctatcct taataaatcc tccactctct ggaaggagac	2700
tgaggggctt tgtaaaacat tagtcagttg ctcattttta tgggattgct tagctgggct	2760
gtaaagatga aggcatacaa taaactcaaa gtatttttaa atttttttga taatagagaa	2820
acttcgctaa ccaactgttc tttcttgagt gtatagcccc atcttggtgt aacttgctgc	2880
ttctgcactt catatccata tttcctattg ttcactttat tctgtagagc agcctgccaa	2940
gaattttatt tctgctgttt tttttgctgc taaagaaagg aactaagtca ggatgttaac	3000
agaaaagtcc acataaccct agaattctta gtcaaggaat aattcaagtc agcctagaga	3060
ccatgttgac tttcctcatg tgtttcctta tgactcagta agttggcaag gtcctgactt	3120
tagtcttaat aaaacattga attgtagtaa aggtttttgc aataaaaaact tactttgg	3178

<210> 44

<211> 1858

<212> DNA

<213> Mus sp.

<400> 44

attccggagg tgaaaaacaa tggcacaacg tgtataatgg ccagcttctc tgcctccttt	60
ctgaccacct acgagactgc gaatggttct cagatcgtga acatttccct gccagcctct	120
gcagaagtac tgaaaaatgg cagttcttgt ggtaaagaaa atgtttctga cccagcctc	180
acaattactt ttggaagagg atatttactg acactcaact tcacaaaaaa tacaacacgt	240
tacagtgtcc agcatatgta ttttacatat aacttgtcag atacagaaca ttttcccaat	300
gccatcagca aagagatcta caccatggat tccacaactg acatcaaggc agacatcaac	360
aaagcatacc ggtgtgtcag tgatatccgg gtctacatga agaattgtgac cgttgtgctc	420
cgggatgcc a ctatccaggc ctacctgtcg agtggcaact tcagcaagga agagacacac	480
tgcacacagg atggaccttc cccaaccact gggccacca gccctcacc accacttgtg	540
cccacaaacc cactgtatc caagtacaat gttactggta acaacggaac ctgcctgctg	600
gcctctatgg cactgcaact gaatatcacc tacctgaaaa aggacaacaa gacggtgacc	660
agagcgttca acatcagccc aaatgacaca tctagtggga gttgcggtat caacttggtg	720
accctgaaag tggagaacaa gaacagagcc ctggaattgc agtttgggat gaatgccagc	780
tctagcctgt ttttcttgca aggagtgcgc ttgaatatga ctcttcctga tgccctagt	840
cccacattca gcatctcaa ccattcactg aaagctcttc aggccactgt gggaaactca	900
tacaagtgca acactgagga acacatcttt gtcagcaaga tgctctccct caatgtcttc	960
agtgtgcagg tccaggcttt caaggtggac agtgacaggt ttgggtctgt ggaagagtgt	1020
gttcaggatg gtaacaacat gttgatcccc attgctgtgg gcggtgccct ggcagggctg	1080
atcctcatcg tcctcattgc ctacctcatt ggcaggaaga ggagtcacgc cggctatcag	1140
accatctagc ctggtgggca ggtgcaccag agatgcacag gggcctgttc tcacatcccc	1200
aagcttagat aggtgtggaa gggaggcaca ctttctggca aactgtttta aaatctgctt	1260
tatcaaatgt gaagttcatc ttgcaacatt tactatgcac aaaggaataa ctattgaaat	1320
gacggtgтта attttgctaa ctgggttaaa tattgatgag aaggctccac tgatttgact	1380
tttaagactt ggtgttttgt tcttcattct tttactcaga tttaagccta tcaaagggat	1440
actctggtcc agaccttggc ctggcaaggg tggctgatgg ttaggctgca cacacttaag	1500
aagcaacggg agcaggaag gcttgcacac aggcacgcac agggtaacc tctggacact	1560
tggcttgggc tacctggcct tgggggggct gaactctggc atctggctgg gtacacaccc	1620

ccccaatttc tgtgctctgc caccctgag ctgccacttt cctaaataga aaatggcatt	1680
atTTTTatTTT actTTTTtTgt aaagtgattt ccagtcttTgt gttggcgTtc agggTggccc	1740
Tgtctctgca ctgtgtacaa taatagattc acactgctga cgtgtctTgc agcgtaggtg	1800
ggtTgtacac tgggcatcag ctcacgtaat gcattgcctg taacgatgct aataaaaa	1858

<210> 45
 <211> 2339
 <212> DNA
 <213> Homo sapiens

<400> 45	
ggcccaaccg ccgcccgcgc ccccgtctc cgcaccgtac ccggccgcct cgcgccatgg	60
cggcccccg cagcgcccg cgacccctgc tgctgctact gctgttgctg ctgctcggcc	120
tcatgcattg tgcgtcagca gcaatgttta tggTgaaaaa tggcaacggg accgcgtgca	180
taatggccaa cttctctgct gccttctcag tgaactacga caccaagagt ggccctaaga	240
acatgacctt tgacctgcca tcagatgcca cagtggTgct caaccgcagc tcctgtggaa	300
aagagaacac ttctgacccc agtctcgtga ttgcttttgg aagaggacat acactcactc	360
tcaatttcac gagaaatgca acacgttaca gcgtccagct catgagtttt gttataact	420
Tgtcagacac acaccttttc cccaatgcga gctccaaaga aatcaagact gtggaatcta	480
taactgacat cagggcagat atagataaaa aatacagatg tgTtagtggc acccaggtcc	540
acatgaacaa cgtgaccgta acgctccatg atgccaccat ccaggcgTac ctttccaaca	600
gcagcttcag caggggagag acacgctgtg aacaagacag gccttcccca accacagcgc	660
cccctgcgcc acccagcccc tcgccctcac ccgtgccc aa gagcccctct gtggacaagt	720
acaacgtgag cggcaccaac gggacctgcc tgctggccag catggggctg cagctgaacc	780
tcacctatga gaggaaggac aacacgacgg tgacaaggct tctcaacatc aaccccaaca	840
agacctcggc cagcgggagc tgcggcgccc acctggTgac tctggagctg cacagcgagg	900
gcaccaccgt cctgctcttc cagttcggga tgaatgcaag ttctagccgg tttttcctac	960
aaggaatcca gttgaataca attcttctctg acgccagaga ccctgccttt aaagctgcca	1020
acggctccct gcgagcgctg caggccacag tcggcaattc ctacaagtgc aacgcggagg	1080
agcacgtccg tgtcacgaag gcgttttTcag tcaatatatt caaagtgtgg gtccaggctt	1140
tcaaggtgga aggtggccag tttggctctg tggaggagtg tctgctggac gagaacagca	1200

tgctgatccc catcgctgtg ggtggtgccc tggcggggct ggtcctcatc gtcctcatcg	1260
cctacctcgt cggcaggaag aggagtcacg caggctacca gactatctag cctggtgcac	1320
gcaggcacag cagctgcagg ggcctctgtt cttttctctg ggcttagggg cctgtcgaag	1380
gggaggcaca ctttctggca aacgtttctc aaatctgctt catccaatgt gaagttcatc	1440
ttgcagcatt tactatgcac aacagagtaa ctatcgaat gacggtgtta attttgctaa	1500
ctgggttaaa tattttgcta actggttaaa cattaatatt taccaaagta ggattttgag	1560
ggtgggggtg ctctctctga gggggtgggg gtgccgctgt ctctgagggg tgggggtgcc	1620
gctgtctctg aggggtgggg gtgccgctct ctctgagggg gtgggggtgc cgctttctct	1680
gaggggggtg ggggtgccgt ctctctgagg ggggtggggg gctgctctct ccgaggggtg	1740
gaatgccgt gtctctgagg ggtgggggtg ccgctctaaa ttggctccat atcatttgag	1800
tttagggttc tgggtgtttg tttcttcatt ctttactgca ctgagattta agccttacia	1860
agggaaagcc tctggccgtc acacgtagga cgcataaggg tcactcgtgg tgaggctgac	1920
atgctcacac attacaacag tagagaggga aaatcctaag acagagggaac tccagagatg	1980
agtgtctgga gcgcttcagt tcagctttaa aggccaggac gggccacacg tggctggcgg	2040
cctcgttcca gtggcggcac gtccttgggc gtctctaag tctgcagctc aagggtggc	2100
acttttttaa atataaaaat ggggtgttatt tttatTTTTT tttgtaaaagt gatttttggg	2160
cttctgttga cattcggggg gatcctgttc tgcgctgtgt acaatgtgag atcgggtgcgt	2220
tctcctgatg ttttgccgtg gcttggggat tgtacacggg accagctcac gtaatgcatt	2280
gcctgtaaca atgtaataaa aagcctcttt cttttaaaaa aaaaaaaaaa aaaaaaaaaa	2339

<210> 46

<211> 45

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 46

cagtacatca aggccaacag caagttcatc ggcataccg aactc

45

<210> 47

<211> 15
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic peptide

<400> 47
Gln Tyr Ile Lys Ala Asn Ser Lys Phe Ile Gly Ile Thr Glu Leu
1 5 10 15

<210> 48
<211> 39
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 48
gctaaatttg tggctgcctg gacactgaaa gccgccgct 39

<210> 49
<211> 13
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic peptide

<400> 49
Ala Lys Phe Val Ala Ala Trp Thr Leu Lys Ala Ala Ala
1 5 10

<210> 50
<211> 593
<212> DNA
<213> Woodchuck hepatitis virus

<400> 50
aatcaacctc tggattacaa aatttgtgaa agattgactg gtattcttaa ctatgttgct 60
ccttttacgc tatgtggata cgctgcttta atgcctttgt atcatgctat tgcttcccgt 120
atggctttca ttttctctc cttgtataaa tcctggttgc tgtctcttta tgaggagttg 180
tggcccgttg tcaggcaacg tggcgtggtg tgcactgtgt ttgctgacgc aacccccact 240

ggttggggca ttgccaccac ctgtcagctc ctttccggga ctttcgcttt cccctccct	300
attgccacgg cggaactcat cgccgcctgc cttgcccgt gctggacagg ggctcggctg	360
ttgggcactg acaattccgt ggtgttgtcg gggaagctga cgtcctttcc atggctgctc	420
gcctgtgttg ccacctggat tctgcgcggg acgtccttct gctacgtccc ttcggccctc	480
aatccagcgg accttccttc ccgcggcctg ctgccggctc tgcggcctct tccgcgtctt	540
cgccttcgcc ctacagacgag tcggatctcc ctttgggccc cctccccgcc tgt	593

<210> 51
 <211> 589
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 51	
tctccccccc cccctctcc ctccccccc cctaagctta ctggccgaag ccgcttgga	60
taaggccggt gtgcgtttgt ctatatgtta tttccacca tattgccgtc ttttggcaat	120
gtgaggggccc ggaaacctgg ccctgtcttc ttgacgagca ttcctagggg tctttcccct	180
ctgcctaaaag gaatgcaagg tctgttgaat gtcgtgaagg aagcagttcc tctggaagct	240
tcttgaagac aaacaacgtc tgtagcgacc ctttgcaggc agcgggaaccc cccacctggc	300
gacaggtgcc tctgcggcca aaagccacgt gtataagata cacctgcaaa ggcggcacaa	360
ccccagtgcc acgttgtgag ttggatagtt gtggaaagag tcaaattggct ctctcaagc	420
gtattcaaca aggggctgaa ggatgccag aaggtacccc attgtatggg atctgatctg	480
gggcctcggg gcacatgctt tacatgtgtt tagtcgaggt taaaaaacg tctaggcccc	540
ccgaaccacg gggacgtggt tttcctttga aaaacacgat gataatatg	589

<210> 52
 <211> 720
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 52

atggtgagca agggcgagga gctgttcacc ggggtggtgc ccatcctggt cgagctggac	60
ggcgacgtaa acggccacaa gttcagcgtg tccggcgagg gcgagggcga tgccacctac	120
ggcaagctga ccctgaagtt catctgcacc accggcaagc tgcccgtgcc ctggcccacc	180
ctcgtgacca ccctgaccta cggcgtgcag tgcttcagcc gctaccccga ccacatgaag	240
cagcacgact tcttcaagtc cgccatgccc gaaggctacg tccaggagcg caccatcttc	300
ttcaaggacg acggcaacta caagaccgc gccgaggtga agttcgaggg cgacaccctg	360
gtgaaccgca tcgagctgaa gggcatcgac ttcaaggagg acggcaacat cctggggcac	420
aagctggagt acaactacaa cagccacaac gtctatatca tggccgacaa gcagaagaac	480
ggcatcaagg tgaacttcaa gatccgccac aacatcgagg acggcagcgt gcagctcgcc	540
gaccactacc agcagaacac ccccatcggc gacggccccg tgctgctgcc cgacaaccac	600
tacctgagca cccagtccgc cctgagcaaa gaccccaacg agaagcgcga tcacatggtc	660
ctgctggagt tcgtgaccgc cgccgggatc actctcggca tggacgagct gtacaagtag	720

<210> 53

<211> 1563

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic
polynucleotide

<400> 53

atgctgctgc tgctgctgct gctgggcctg aggctacagc tctccctggg catcatcca	60
gttgaggagg agaaccgga cttctggaac cgcgaggcag ccgaggccct gggtgccgcc	120
aagaagctgc agcctgcaca gacagccgcc aagaacctca tcatcttcct gggcgatggg	180
atgggggtgt ctacggtgac agctgccagg atcctaaaag ggcagaagaa ggacaaactg	240
gggcctgaga taccctggc catggaccgc ttcccatatg tggctctgtc caagacatac	300
aatgtagaca aacatgtgcc agacagtgga gccacagcca cggcctacct gtgcggggtc	360
aagggaact tccagacat tggcttgagt gcagccgcc gctttaacca gtgcaacacg	420
acacgaggca acgaggtcat ctccgtgatg aatcgggcca agaaagcagg gaagtcagtg	480
ggagtggtaa ccaccacacg agtgcagcac gcctcgccag ccggcaccta cgccacacg	540
gtgaaccgca actggtactc ggacgccgac gtgcctgcct cggcccgcga ggaggggtgc	600

caggacatcg ctacgcagct catctccaac atggacattg acgtgatcct aggtggaggc	660
cgaagtaga tgtttcgcat gggaaccca gaccctgagt acccagatga ctacagccaa	720
ggtgggacca ggctggacgg gaagaatctg gtgcaggaat ggctggcgaa gcgccagggt	780
gcccggtatg tgtggaaccg cactgagctc atgcaggctt ccctggaccc gtctgtgacc	840
catctcatgg gtctctttga gcctggagac atgaaatacg agatccaccg agactccaca	900
ctggaccctt ccctgatgga gatgacagag gctgccctgc gcctgctgag caggaacccc	960
cgcggttctt tcctcttcgt ggagggtggt cgcatcgacc atggtcatca tgaaagcagg	1020
gcttaccggg cactgactga gacgatcatg ttcgacgacg ccattgagag ggcgggccag	1080
ctcaccagcg aggaggacac gctgagcctc gtcactgccg accactccca cgtcttctcc	1140
ttcggagggt accccctgcg agggagctcc atcttcgggc tggcccctgg caaggcccgg	1200
gacaggaagg cctacacggt cctcctatac ggaaacggtc caggctatgt gctcaaggac	1260
ggcgcccggc cggatgttac cgagagcgag agcgggagcc ccgagtatcg gcagcagtca	1320
gcagtgcccc tggacgaaga gaccacgca ggcgaggacg tggcgggtgtt cgcgcgcggc	1380
ccgcaggcgc acctggttca cggcgtgcag gagcagacct tcatagcgca cgtcatggcc	1440
ttcgccgcct gcctggagcc ctacaccgcc tgcgacctgg cgcccccgcc cggcaccacc	1500
gacgccgcgc acccgggtta ctctagagtc ggggcggccg gccgcttcga gcagacatga	1560
taa	1563

<210> 54

<211> 1653

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 54

atggaagatg ccaaaaacat taagaagggc ccagcgccat tctaccact cgaagacggg	60
accgccggcg agcagctgca caaagccatg aagcgctacg ccctggtgcc cggcaccatc	120
gcctttaccg acgcacatat cgagggtggac attacctacg ccgagtactt cgagatgagc	180
gttcggctgg cagaagctat gaagcgctat gggctgaata caaacatcg gatcgtggtg	240

tgcagcgaga atagcttgca gttcttcatg cccgtgttgg gtgccctgtt catcggtgtg	300
gctgtggccc cagctaacga catctacaac gagcgcgagc tgctgaacag catgggcatc	360
agccagccca ccgtcgtatt cgtgagcaag aaagggtgc aaaagatcct caacgtgcaa	420
aagaagctac cgatcataca aaagatcatc atcatggata gcaagaccga ctaccagggc	480
ttccaaagca tgtacacctt cgtgacttcc catttgccac ccggcttcaa cgagtacgac	540
ttcgtgcccg agagcttcga ccgggacaaa accatcgccc tgatcatgaa cagtagtggc	600
agtaccggat tgcccaaggg cgtagcccta ccgcaccgca ccgcttgtgt ccgattcagt	660
catgcccgcg accccatctt cggcaaccag atcatccccg acaccgctat cctcagcgtg	720
gtgccatttc accacggctt cggcatgttc accacgtgg gctacttgat ctgaggcttt	780
cgggtcgtgc tcatgtaccg cttcaggag gagctattct tgcgcagctt gcaagactat	840
aagattcaat ctgccctgct ggtgccaca ctatttagct tcttcgctaa gagcactctc	900
atcgacaagt acgacctaag caacttgac gagatcgcca gcggcggggc gccgctcagc	960
aaggaggtag gtgaggccgt ggccaaacgc ttccacctac caggcatccg ccagggtac	1020
ggcctgacag aaacaaccag cgccattctg atcacccccg aaggggacga caagcctggc	1080
gcagtaggca aggtggtgcc cttcttcgag gctaagggtg tggacttgga caccggtaag	1140
acactgggtg tgaaccagcg cggcgagctg tgcgtccgtg gccccatgat catgagcggc	1200
tacgttaaca accccgaggc tacaaacgct ctcatcgaca aggacggctg gctgcacagc	1260
ggcgacatcg cctactggga cgaggacgag cacttcttca tcgtggaccg gctgaagagc	1320
ctgatcaaat acaagggtta ccaggtagcc ccagccgaac tggagagcat cctgctgcaa	1380
cacccaaca tcttcgacgc cggggtcgcc ggcctgcccg acgacgatgc cggcgagctg	1440
cccgccgcag tcgtcgtgct ggaacacggt aaaaccatga ccgagaagga gatcgtggac	1500
tatgtggcca gccaggttac aaccgccaag aagctgcgcg gtggtgttgt gttcgtggac	1560
gaggtgccta aaggactgac cggcaagttg gacgcccga agatccgcga gattctcatt	1620
aaggccaaga agggcggcaa gatcgccgtg taa	1653

<210> 55

<211> 66

<212> DNA

<213> Foot-and-mouth disease virus

<400> 55
gtaaagcaaa cactgaactt tgaccttctc aagttggctg gagacgttga gtccaatcct 60
gggccc 66

<210> 56
<211> 5
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
peptide

<400> 56
Gly Pro Gly Pro Gly
1 5